

UNIVERZA V LJUBLJANI  
FAKULTETA ZA RAČUNALNIŠTVO IN INFORMATIKO

Dušan Kalanj

**Metode prepoznavanja obrazov ob  
nenadzorovanih pogojih zajema slik**

MAGISTRSKO DELO

ŠTUDIJSKI PROGRAM DRUGE STOPNJE  
RAČUNALNIŠTVO IN INFORMATIKA

MENTOR: prof. dr. Franc Solina

Ljubljana, 2016

Rezultati magistrskega dela so intelektualna lastnina avtorja in Fakultete za računalništvo in informatiko Univerze v Ljubljani. Za objavljanje ali izkoriščanje rezultatov magistrskega dela je potrebno pisno soglasje avtorja, Fakultete za računalništvo in informatiko ter mentorja.

*Rad bi se iskreno zahvalil prof. dr. Francu Solini za mentorstvo in potrpežljivost, prof. dr. Grgiću za dovoljenje pri uporabi baze obrazov SCface, dr. Nguyenu za pojasnila ter viš. pred. dr. Borutu Batagelju za strokovno vodenje, nasvete in nesebično pomoč.*





For Frodo.



# Kazalo

Povzetek

Abstract

<b>1</b>	<b>Uvod</b>	<b>1</b>
1.1	Kontekst . . . . .	1
1.2	Biometrični sistemi . . . . .	2
1.3	Sistemi za prepoznavanje obrazov . . . . .	3
1.4	Prispevki dela . . . . .	11
1.5	Struktura magistrskega dela . . . . .	12
<b>2</b>	<b>Baze obrazov</b>	<b>13</b>
2.1	FERET . . . . .	13
2.2	SCface . . . . .	15
2.3	ChokePoint . . . . .	19
<b>3</b>	<b>Metode prepoznavanja obrazov</b>	<b>23</b>
3.1	Detekcija obrazov . . . . .	23
3.2	Normalizacija . . . . .	25
3.3	Ekstrakcija značilnk . . . . .	35
3.4	Zmanjšanje dimenzionalnosti . . . . .	46
3.5	Klasifikacija . . . . .	50
<b>4</b>	<b>Analiza rezultatov</b>	<b>53</b>
4.1	Rezultati testiranja na bazi FERET . . . . .	53

4.2	Rezultati testiranja na bazi SCface . . . . .	60
4.3	Rezultati testiranja na bazi ChokePoint . . . . .	68
4.4	Nastavitev parametrov . . . . .	70
4.5	Časovna zahtevnost . . . . .	73
4.6	Povzetek analize rezultatov . . . . .	75
<b>5</b>	<b>Zaključek</b>	<b>77</b>
	<b>Literatura</b>	<b>81</b>

# Seznam uporabljenih kratic

kratica	angleško	slovensko
<b>ELBP</b>	elliptical local binary patterns	eliptični lokalni binarni vzorci
<b>KNN</b>	k-nearest neighbor	metoda k-najbližjih sosedov
<b>LBP</b>	local binary patterns	lokalni binarni vzorci
<b>LDA</b>	linear discriminant analysis	analiza linearne diskriminante
<b>LPOG</b>	local patterns of gradients	lokalni vzorci gradientov
<b>LPQ</b>	local phase quantization	kvantizacija lokalne faze
<b>PCA</b>	principal component analysis	metoda glavnih komponent
<b>SVM</b>	support vector machine	metoda podpornih vektorjev



# Povzetek

**Naslov:** Metode prepoznavanja obrazov ob nenadzorovanih pogojih zajema slik

**Avtor:** Dušan Kalanj

V okviru magistrskega dela smo preiskovali prepoznavanje obrazov pri nenadzorovanih pogojih, kamor prištevamo spremenljivo pozo in osvetlitev, obrazni izraz, nizko ločljivost, staranje, prekrivanje ter infrardeče slike. Delo tako vsebuje pregled področja ter temeljit opis metodologije, uporabljene pri prototipu sistema za prepoznavanje obrazov, ki je nastal v procesu. Pri tem smo se osredotočili na korak ekstrakcije značilk, preučili pa smo tudi različne rešitve v korakih normalizacije, zmanjševanja dimenzionalnosti ter klasifikacije. Delo se zaključi z analizo doseženih rezultatov pri testih na bazah obrazov FERET, SCface in ChokePoint, kjer smo preučili tudi efekt frontalizacije. Pri bazah SCface in ChokePoint smo, vsaj kolikor nam je znano, dosegli najboljše rezultate v literaturi.

**Ključne besede:** Prepoznavanje obrazov, nenadzorovani pogoji, ekstrakcija značilk, normalizacija, zmanjševanje dimenzionalnosti, FERET, SCface, ChokePoint.





# Abstract

**Title:** Face recognition methods under uncontrolled conditions of image capture

**Author:** Dušan Kalanj

In the scope of this master thesis we investigated the problem of face recognition in unconstrained environments, where a system has to be robust against varying face pose, expression and illumination as well as low resolution, ageing of subjects and infrared capturing mode. The work contains a survey of the field and a detailed description of the methodology used in the making of the face recognition system prototype that was developed in the process. While focusing on feature extraction, we also examined different solutions in the stages of normalization, dimensionality reduction and classification. The work concludes with an analysis of the results achieved during testing, which was performed on the FERET, SCface and ChokePoint datasets. On the first two of the three mentioned datasets we also examined the effect of frontalization. To the best of our knowledge, we achieved the best results in literature on SCface and ChokePoint datasets.

**Keywords:** Face recognition, uncontrolled conditions, feature extraction, normalization, dimensionality reduction, FERET, SCface, ChokePoint.



# Poglavje 1

## Uvod

To poglavje je namenjeno seznanitvi z osnovnimi koncepti tehnologije prepoznavanja obrazov, njeni uporabi in problematiki, s katero se področje trenutno sooča. Poglavje vsebuje tudi predstavitev ciljev in prispevkov teze ter se zaključi s povzetkom preostalih poglavij.

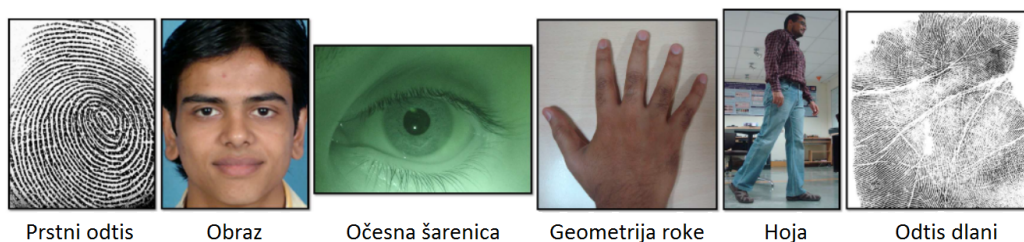
### 1.1 Kontekst

Kamorkoli se danes ozremo, vidimo obraze. Naj bo to v živo, na namizni fotografiji, v časopisu, na obcestnem panoju ali pa na zaslonu. V zadnjih desetletjih smo priča obsežni disipaciji na slednjem mediju, kar gre pripisati zlasti razvoju ter dostopnosti naprav za zajem slik in pojavu socialnih omrežij, oboje pa je nedvomno posledica digitalne revolucije.

Vzporedno z eksponentnim porastom obrazov na zaslonih pa so se povečale tudi priložnosti za uporabo na novo pridobljenih informacij, kar je vodilo v sunkovit razvoj tehnologije za prepoznavanje obrazov. Tako se je njena uporaba v zadnjih dveh desetletjih razširila na številna področja, med katerimi na prvih mestih najdemo video nadzor, forenziko, komunikacijo človek računalnik (predvsem za namen avtentikacije), obrazni tagging, itd. A da bi do potankosti doumeli obseg njihove uporabe ter težave, s katerimi se soočajo, je najprej potrebno razložiti, kako delujejo.

## 1.2 Biometrični sistemi

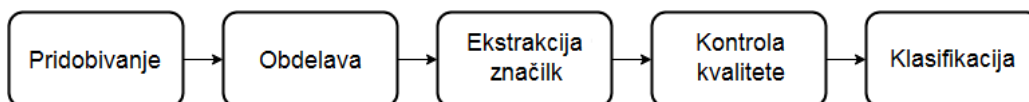
Sistemi za prepoznavanje obrazov se uvrščajo med biometrične sisteme, kamor spadajo tudi na primer sistemi za identifikacijo na podlagi očesne šarenice, mrežnice, prstnega odtisa in glasu [7]. Primere takšnih biometričnih lastnosti lahko opazujemo na sliki 1.1.



Slika 1.1: Primeri biometričnih značilnosti [10].

Vsi omenjeni sistemi sledijo isti koračni shemi, prikazani na diagramu na sliki 1.2. V prvem koraku se izvede pridobivanje določene biometrične značilnosti osebe. Primer takšnega procesa je detekcija obraza. Temu sledi obdelava pridobljenih podatkov, kot naslednja pa se nad obdelanimi podatki uporabi določena metoda za ekstrakcijo značilnk, ki iz podatkov izlušči najbolj razpoznavne lastnosti.

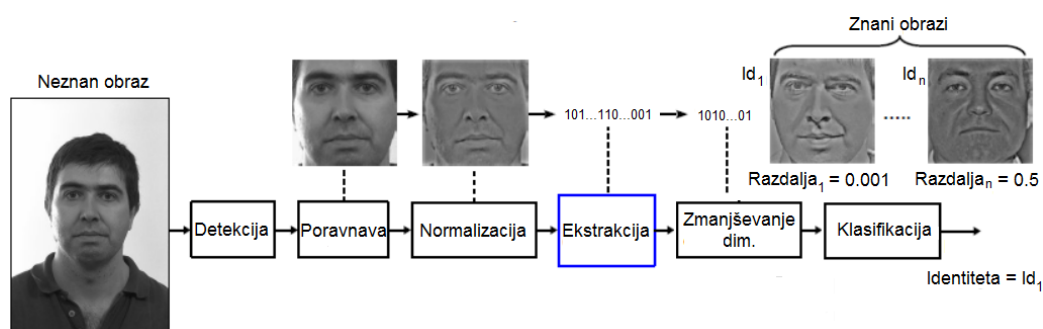
Za tem se izvede t.i. preverjanje kvalitete, kjer se določene pridobljene značilke zavrže, v kolikor se ugotovi, da prispevajo premalo k prepoznavnosti osebe. Proces se zaključi s korakom klasifikacije, kjer se značilke zadevne osebe primerjajo z vnaprej shranjenimi referenčnimi primeri, ki so predhodno pridobljeni po istem postopku. Na ta način se osebi dodeli identiteta [49].



Slika 1.2: Diagram konceptualnega biometričnega sistema [49].

### 1.3 Sistemi za prepoznavanje obrazov

Kot že rečeno, sisteme za prepoznavanje obrazov uvrščamo med biometrične sisteme, ki kot taki sledijo zgoraj opisani koračni shemi. Podrobneje opisano shemo, ki ponazarja korake sistema za prepoznavanje obrazov, lahko opazujemo na sliki 1.3. Pod sliko sledi še podroben opis korakov.



Slika 1.3: Diagram sistema za prepoznavanje obrazov [23].

#### Detekcija obraza

Pri dani sliki je naloga detekcije obraza najti področje slike, ki vsebuje obraz, ne glede na njegovo lokacijo, rotacijo in svetlobne pogoje. Rezultat tega koraka je slika, ki vsebuje le področje obraza. V literaturi sta najbolj popularni rešitvi za detekcijo obrazov Rowley [46] ter Viola in Jones [62].

#### Obdelava slike oz. normalizacija

Naloga tega koraka je minimizirati vpliv lastnosti slik, ki so spremenljive glede na obraz in mu ne pripadajo intrinzično. Pod omenjene lastnosti se prišteva osvetlitev, ločljivost slike ter rotacija obraza. Rezultat tega koraka je normalizirana slika obraza. Zelo popularna rešitev za normalizacijo osvetlitve slik je delo od Tana in Triggsa [55], rotiranje, skaliranje ter obrazovanje pa so trivialne operacije, ki se ne razlikujejo glede na različna dela.

## Ekstrakcija značilnk

V tem koraku se nad obdelano sliko uporabi izbrana metoda za ekstrakcijo značilnk, katere naloga je, da iz slike zajame najbolj razpoznavne lastnosti obraza za namen čim učinkovitejše klasifikacije. Rezultat tega koraka je t.i. vektor značilnk, ki je unikaten za vsak obraz. Metode se v literaturi [72] v glavnem delijo na globalne in lokalne, najbolj znani predstavnici vsake družine pa sta Eigenfaces [61] ter Lokalni binarni vzorci (ang. *Local Binary Patterns* - LBP) [2].

## Zmanjšanje dimenzionalnosti

Da bi bili pridobljeni vektorji značilnk kar se da diskriminantni, se v naslednjem koraku uporabi določena metoda s področja strojnega učenja. Omenjena metoda zmanjša dimenzionalnost vektorjev tako, da ohrani le  $n$  število najbolj diskriminantnih značilnk. Najbolj znani metodi, ki opravljajo opisano nalogo, sta Metoda glavnih komponent (ang. *Principal Component Analysis* - PCA) [19] ter Analiza linearne diskriminante (ang. *Linear Discriminant Analysis* - LDA) [56].

## Klasifikacija

Pri sistemih za prepoznavanje obrazov razpolagamo z dvema vrstama obraznih slik: slike znanih oseb oz. označene ali referenčne slike ter slike neznanih oseb. V zadnjem koraku se torej neznanemu obrazu dodeli identiteta, ki pa se podeli glede na njegovo stopnjo podobnosti z vsemi referenčnimi, predhodno shranjenimi obrazi, ki so bili pridobljeni po istem postopku. V literaturi sta največkrat omenjeni rešitvi Metoda podpornih vektorjev (ang. *Support Vector Machine* - SVM) [22] ter metoda k-najbližjih sosedov (ang. *k-Nearest Neighbor* - KNN) [40].

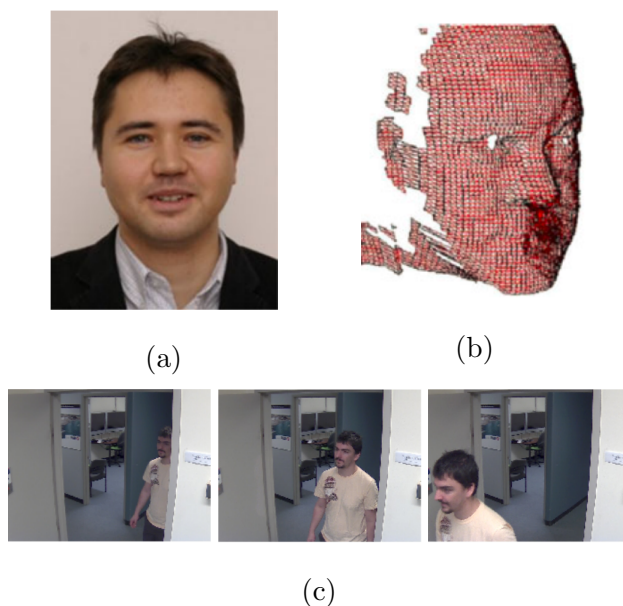
Zgoraj so opisani koraki procesa obrazne identifikacije, sistemi za obrazno prepoznavanje pa znajo opravljati še eno nalogo in sicer obrazno verifikacijo. Slednja se razlikuje v tem, da ima neznana oseba pri vstopu v sistem že

dodeljeno domnevno identiteto, ki se nato v koraku klasifikacije preverja s samo enim znanim obrazom. Delo, predstavljeno v tej tezi, se nanaša samo na obrazno identifikacijo, a bomo ohranili poimenovanje prepoznavanje obrazov.

Posebno opombo si zasluži še opis prvega koraka, kjer smo do sedaj predvidevali, da v sistem vselej posreduje 2D sliko. Sistem je lahko namreč prirejen tudi za video posnetke, sekvence slik ter 3D slike, primere katerih lahko opazujemo na sliki 1.4, a v tem delu smo se osredotočili le na 2D slike zaradi sledečih razlogov:

- Tudi pri video in slikovnih sekvencah je elementarna podatkovna enota, ki se v danem trenutku procesira, še vedno 2D slika.
- Delo s 3D slikami zahteva naprednejša orodja ter veliko večjo računsko moč kakor delo z 2D slikami, s čimer pa pri našem delu žal nismo razpolagali.
- Naše delo se posveča predvsem prepoznavanju obrazov pri nenadzorovanih pogojih. To je, ko sistem razpolaga le z delnim sodelovanjem zadevne osebe. Pri takšnih pogojih je 3D rekonstrukcija obraza zelo otežena.

Če torej povzamemo delovanje sistemov za prepoznavanje obrazov, je rešitev na videz zelo enostavna: ko sistemu predstavimo sliko neznane osebe, bo sliko obdelal, izluščil najbolj prepoznavne značilnosti ter jih primerjal s tistimi od slik znanih oseb. Tako bo identiteta najbolj podobne znane osebe postala identiteta neznane osebe. Opisan princip je zelo podoben človeškemu mehanizmu za prepoznavo: ko vidimo obraz, bomo zajeli njegove najbolj prepoznavne značilnosti kot so oči, usta, nos, ušesa, oblika obraza in barva kože ter v svojem spominu poiskali ime osebe, ki opisu najbolj ustreza. Kot taki so se sistemi za prepoznavanje obrazov razširili na številna področja, ki so opisana v naslednjem poglavju.



Slika 1.4: (a) 2D slika (b) 3D slika (c) Sekvenca slik

### 1.3.1 Uporaba

Sistemi za prepoznavanje obrazov se po učinkovitosti že postavljajo ob bok sistemom za prepoznavanje na podlagi drugih biometričnih lastnosti. A za razliko od slednjih, ki zahtevajo popolno sodelovanje osebe preko fizičnega kontakta ali pozornosti, kot to počnejo sistemi za prepoznavanje na podlagi očesne šarenice ali prstnega odtisa, lahko sistemi za prepoznavanje obrazov pridobijo sliko osebe brez njihovega sodelovanja. Kar naredi prepoznavanje obrazov še bolj zanimivo je dejstvo, da vsebujejo slike obrazov še veliko drugih uporabnih informacij kot so spol, izraz, starost, etničnost in smer pogleda. Napredek na področju digitalnih fotoaparátov in kamer ter pojav socialnih omrežji in storitev za izmenjavo slik sta še dodatno pospešila razvoj na področju prepoznavanja obrazov. Ne čudi torej dejstvo, da so se v zadnjih dveh desetletjih razvili številni načini uporabe tehnologije prepoznavanja obrazov, ki jih lahko v grobem razdelimo na štiri različna področja in sicer na nadzor dostopa, varnost, preiskavo baz obrazov ter ciljno oglaševanje [39]. Sledi opis naštetih področij.



## Nadzor dostopa

Sistemi za prepoznavanje obrazov identificirajo osebe glede na slike njihovih obrazov. Za razliko od običajnih sistemov za avtorizacijo, kot je na primer preverjanje ključa ali digitalne kartice, so ti sistemi zmožni ugotoviti prisotnost dejanske osebe. Takšni sistemi so običajno zelo učinkoviti, saj vsebujejo relativno majhne množice znanih obrazov, zato je sistemu med njimi enostavno razlikovati. Primer takšnega sistema je odklepanje računalnika ob prisotnosti avtorizirane osebe.

## Varnost

Danes je varnost na javnih mestih bolj pomembna kot kdajkoli prej, posledica česar je razvoj številnih varnostnih sistemov, ki se poslužujejo tehnologije prepoznavanja obrazov. Takšen primer je rešitev na letališču Fresno Yosemite International (FYI) v Kaliforniji, kjer so oktobra 2001 implementirali sistem, ki pri prehodu osebe skozi varnostno kontrolo opozori letališko osebje v primeru, da je posameznik podoben znanemu teroristu ali kriminalcu. Primer varnostnega posnetka iz letališča lahko opazujemo na sliki 1.5.



Slika 1.5: Slika osumljencev terorističnega napada na bruseljskem letališču 22. marca 2016, kot jo je posnela varnostna kamera [12].

### **Preiskava baz obrazov**

Sem prištevamo iskanje posameznikov v bazah registriranih voznikov, izgubljenih otrok, imigrantov, policijskih kartotek itd.

### **Ciljno oglaševanje**

Sistemi za ciljno oglaševanje prilagajajo oglase glede na določeno lastnost osebe. Primer takšnega sistema so implementirali na avtobusnih postajah v Združenem Kraljestvu leta 2012 v okviru kampanje "Because I'm a Girl". Sistem je bil zmožen razlikovati med spoloma ter ustrezno prilagajati oglas, čigar namen je bil ozaveščati prebivalstvo o enakopravnosti med spoloma. Leta 2013 je bila v Laboratoriju za računalniški vid tudi na Fakulteti za računalništvo in informatiko v Ljubljani opravljena študija, kjer se je na podlagi obrazov preizkovala demografika strank trgovine z oblekami [45]. S takšnimi informacijami je moč v skoraj realnem času prilagajati vsebino v izlozbi ustrezno s spreminjanjem demografike strank tekom dneva.

Zgoraj opisani sistemi se danes že približujejo 100% učinkovitosti [34, 9], a le, ko oseba z njimi vsaj delno sodeluje. V nasprotnem primeru učinkovitost drastično pade [9]. Zakaj je temu tako, bo razloženo v naslednjem razdelku.

## **1.3.2 Problematika**

Večina danes uporabljenih sistemov za prepoznavanje obrazov zahteva od osebe vsaj delno sodelovanje. Primer takšnega sodelovanja je npr. vzporedno poravnanje obraza s kamero na računalniku za namen odklepanja. A če bi želeli isti sistem odklepanja na podlagi obraza prenesti iz osebne računalnika še na vhodna vrata stanovanja, se stvari zakomplicirajo. Kamera bi namreč bila postavljena dlje od obraza, kar bi se poznalo predvsem na ločljivosti, pa tudi na položaju in rotaciji obraza. Prav tako bi na učinkovitost vplivalo vreme, ki bi skozi okno ob različnih dneh različno osvetlilo vaš obraz. Isti vpliv se seveda pozna tudi pri odklepanju računalnika, a pride pri vratih zaradi manjše ločljivosti do večjega izraza. Zagotovo bi

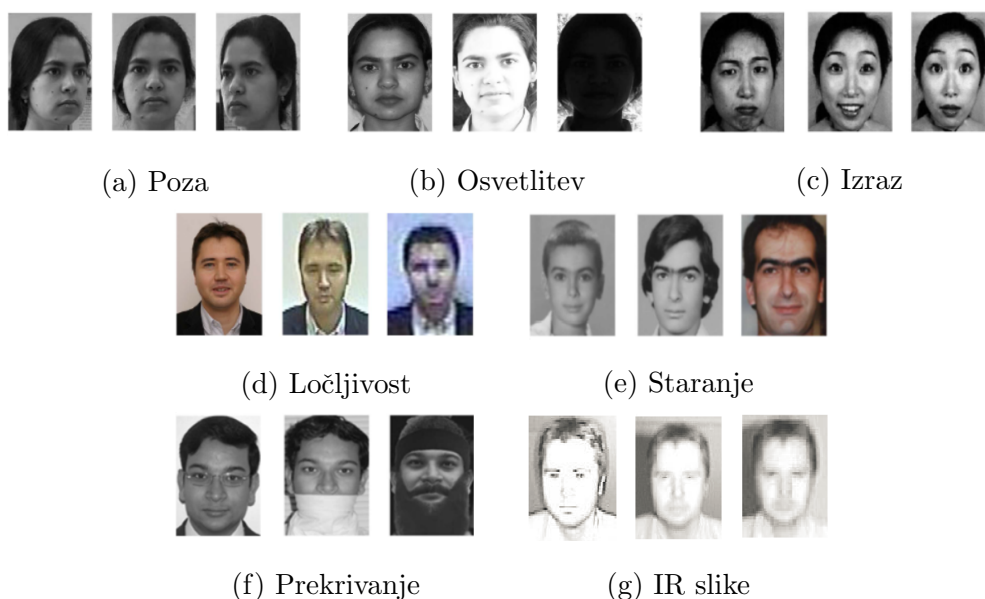
se tudi kdaj vračali s prijatelji, kar bi sistem še dodatno zmedlo. Prišel pa bi tudi dan, ko bi bili resnično nepripravljeni sodelovati s kamero na oboku vrat, pa naj bo to zaradi tega, ker ste se spomnili, da ste pustili prižgano pečico, ali pa vas le muči potreba. Če vemo, da bi vas tak sistem zagotovo kdaj pustil na cedilu, si predstavljajte, kakšne preglavice bi šele imel sistem, ki bi poiskoval identificirati vlomilca, ki s prestrašenim izrazom teče čez Tromostovje, najbližja kamera pa bi se nahajala pri spomeniku dr. Franceta Prešerna.

Zgoraj so navedene le nekatere izmed težav, na katere naletijo sistemi za prepoznavanje obrazov pri nenadzorovanih pogojih. Spodaj so, tokrat bolj strokovno, opisani najbolj pogosti izzivi, s katerimi se sooča področje prepoznavanja obrazov ob delovanju pri nenadzorovanih pogojih [10], moč pa jih je opazovati tudi na sliki 1.6.

- **Spremenljiva poza:** slike obrazov so zelo spremenljive glede na kot, pod katerim so zajete. Če je kot glede na prednji del obraza večji od  $45^\circ$ , učinkovitost sistema drastično pade. To je posledica dejstva, da so obrazi poznanih oseb skoraj vedno slikani frontalno ter, da frontalne slike vsebujejo veliko več informacij o obrazu kakor slike iz profila ali pol-profila.
- **Spremenljiva osvetlitev:** pridobljene slike lahko izkazujejo velike razlike pri osvetlitvi glede na slike poznanih obrazov, ki so ponavadi pridobljene pri enakomerni osvetlitvi. Vpliv se pozna zlasti pri slikah nižje ločljivosti.
- **Obrazni izrazi:** spremembe v izrazu glede na slike poznanih oseb lahko povzročijo deformacijo ključnih obraznih lastnosti kot so oči, obrvi, usta in nos ter s tem vplivajo na učinkovitost sistema.

- **Nizka ločljivost:** omenjena težava je najbolj očitna pri nadzornih kamerah, ki so primarno zasnovane za nadzor večjega območja iz ene lokacije. Zavaljo večje pokritosti, manjše velikosti posnetkov ter operacijskih stroškov je ločljivost slik zmanjšana. Posledica tega je manjša izrazitost obraznih značilnosti, kar vodi v manjšo učinkovitost sistemov.
- **Staranje:** poleg tega, da se s časom spreminjajo obrazne lastnosti kot so lasje, brada in polt, se spreminja tudi struktura obraza. Slednje seveda velja le za dolgoročno staranje in se najbolj očitno spreminja v prvi četrtini življenja.
- **Prekrivanje in preobrazba:** sem prištevamo dodatke, ki niso intrinzični obrazu. Primeri takšnih dodatkov so pokrivala, očala, šali in ličila. Omenjeni dodatki zelo vplivajo na učinkovitost sistema ter predstavljajo enega najtežjih izzivov pri napredku sistemov za prepoznavanje obrazov.
- **Infrardeče slike:** omenjen način uporabljajo predvsem nadzorne kamere pri snemanju ponoči. Poleg tega, da je njihova osvetlitev zelo drugačna tisti na slikah poznanih obrazov, vpliva na infrardeče slike še temperatura okolja.

Če bi se zgoraj opisani problemi pri nenadzorovanih pogojih pojavljali posamezno, bi jih bilo veliko lažje reševati, kakor pa je to primer v resničnem scenariju. Skupaj namreč zelo popačijo podobo posameznika na sliki, posledica česar je, da jih težko prepoznajo tudi ljudje. In če to ni dovolj, je tu še **problem zelo velikih sistemov**. Takšni sistemi lahko vsebujejo tudi po več sto milijonov slik poznanih oseb, kakor je primer pri bazah obrazov prebivalcev večjih držav. Omenjen problem je težko rešiti že, če sistemu predstavimo kvalitetno sliko obraza, v nasprotnem primeru pa skoraj nemogoče. Na vprašanje, kako narediti tak sistem prepoznavanja obrazov učinkovit, je zelo težko odgovoriti.



Slika 1.6: Problemi, s katerimi se soočajo sistemi za prepoznavanje obrazov ob delovanju pri nenadzorovanih pogojih [10].

## 1.4 Prispevki dela

V okviru tega dela smo se lotili reševanja zgoraj opisanih problemov. To smo naredili z implementacijo sistema za prepoznavanje obrazov pri upoštevanju najboljših priporočil iz literature s poudarkom na prepoznavanju pri nenadzorovanih pogojih. Prispevki dela so torej sledeči:

- Opisi metod za normalizacijo obrazov (svetlobna izravnava, frontalizacija), ekstrakcijo značilk, zmanjšanje dimenzionalnosti vektorjev značilk ter klasifikacijo glede na najnovejša priporočila iz literature.
- Analiza učinka različnih rešitev iz zgoraj omenjenih korakov na rezultate prepoznavanja obrazov pri testiranju na bazah FERET [43], SCface [19] in ChokePoint [64].
- Obsežen opis rezultatov na bazah SCface in ChokePoint, ki sta, za razliko od baze FERET zelo slabo raziskani.

## 1.5 Struktura magistrskega dela

V nadaljevanju bomo najprej predstavili baze obrazov, na katerih smo našo rešitev testirali. V tretjem poglavju bo sledil kratek pregled literature z opisom metod, ki smo jih uporabili pri implementaciji sistema. V četrtem poglavju bomo opisali analizo rezultatov testiranja ter delo zaključili z idejami za nadaljno raziskavo ter sklepnimi mislimi.

## Poglavje 2

# Baze obrazov

V tem poglavju bomo opisali baze obrazov, ki smo jih uporabili pri testiranju naše rešitve. Omenjene baze so FERET [43], SCface [19] ter ChokePoint [64], njihovi opisi pa si sledijo v istem zaporedju.

### 2.1 FERET

Do nastanka baze obrazov FERET ni bilo možno natančno oceniti ali primerjati algoritmov za prepoznavanje obrazov. Številni raziskovalci so za probleme, ki so jih preiskovali, sami ustvarjali baze obrazov, ki pa so bile majhne in nekonsistentne. Zato sta se dr. Wechsler in dr. Phillips iz Univerze George Mason v ZDA odločila, da bosta ustvarila obsežno bazo obrazov neodvisno od razvijalcev algoritmov. Rezultat je bila baza FERET (Face Recognition Technology), ki se je na področju prepoznavanja obrazov uveljavila kot najbolj vplivna in največkrat citirana baza obrazov. Sledi natančnejši opis baze.

#### 2.1.1 Pridobivanje slik

Baza FERET je nastala v 15-ih sejah med leti 1993 in 1996. Slike so nastale v srednje nadzorovanem okolju, za ohranitev konsistenčnosti pa se je pri fotografskih sejah venomer uporabljala ista konfiguracija. Baza vsebuje obraze 1,196-ih posameznikov, med katerimi sta za nekatere minili več kot dve leti

med prvim in zadnjim fotografiranjem. Omenjen časovni zamik je zelo pomemben za raziskovalce, saj jim je prvič omogočil raziskovanje dolgoročnih sprememb na videzu posameznikov [43].

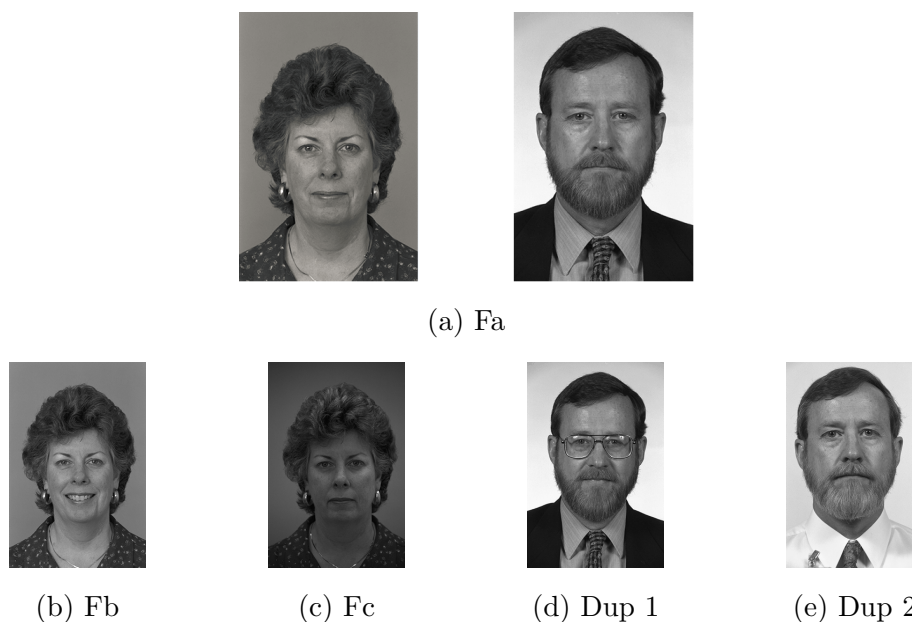
### 2.1.2 Struktura baze in lastnosti slik

Baza vsebuje 14,051 črno-belih slik velikosti  $256 \times 384$  pikslov. Na slikah so obrazi s pogledi, ki segajo od frontalnih do levih in desnih profilov. Skozi leta so se v okviru različnih raziskav oblikovale številne testne množice, a najbolj se je obdržala razdelitev od Phillips et al. [43]. Pri omenjeni razdelitvi obstaja pet množic s poimenovanji Fa, Fb, Fc, Dup 1 ter Dup 2. Množica Fa služi kot zbirka poznanih oz. referenčnih obrazov med tem, ko so ostale namenjene prepoznavanju. V Množici Fb so slike na katerih posamezniki glede na množico Fa izkazujejo največje razlike pri obraznem izrazu, množica Fc pa vsebuje slike z največjimi razlikami pri osvetlitvi. Množica Dup 1 vsebuje slike, ki so bile narejene od ene minute pa do 1031 dni za tistimi v množici Fa, v množici Dup 2 pa so vsebovani najtežji primeri iz množice Dup 1. Tabela 2.1 vsebuje informacije o številu vsebovanih slik ter namenu testiranja posamezne množice, na sliki 2.1 pa lahko opazujemo primere iz vsake izmed množic<sup>9</sup>

Tabela 2.1: Osnovne informacije o bazi FERET [43].

Množica	Število slik	Namen testiranja
Fa	1,196	Znani obrazi
Fb	1,195	Izraz na obrazu
Fc	194	Osvetlitev obraza
Dup 1	722	Staranje oseb
Dup 2	234	Staranje oseb





Slika 2.1: Primeri slik iz baze FERET

## 2.2 SCface

V zadnjih dveh desetletjih so nastale številne baze obrazov za namen testiranja tehnologije prepoznavanja obrazov, žal pa nobena raziskovalcem ni omogočala preverjanja rešitev pri nenadzorovanih pogojih kot so spremenljiva poza, osvetlitev ter ločljivost slike. Dodatno so bile slike pridobljene le z eno napravo, ki je bila običajno boljša od enostavnih, komercialno dostopnih naprav, ki se v resničnem svetu uporabljajo pri video nadzoru. Zgoraj navedena dejstva so predstavljala glavno motivacijo za nastanek baze obrazov SCface (Surveillance Cameras Face Database), za kar so zaslužni Grgić et. al [19] z Univerze v Zagrebu.

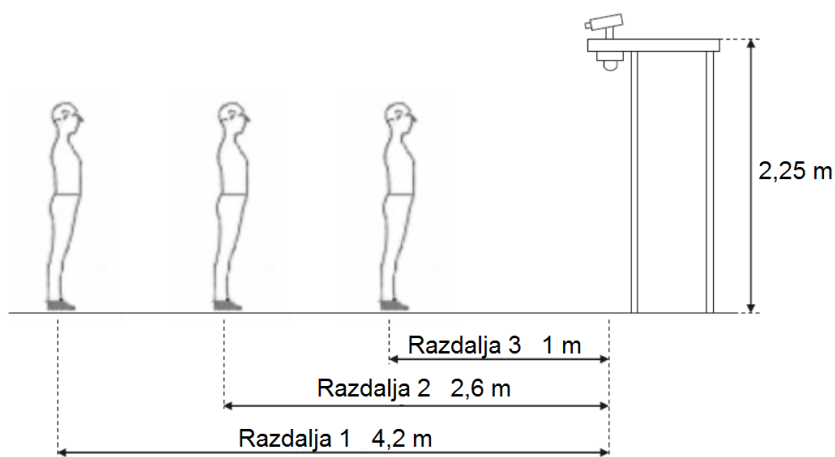
Baza obrazov SCface vsebuje 4,160 slik 130-ih različnih oseb. Slike so bile narejene iz treh različnih razdalj in z uporabo petih različno kakovostnih kamer, ki so namenjene video nadzoru. Baza vsebuje tudi slike, narejene pri delovanju v infrardečem nočnem načinu, česar sta bile sposobne dve izmed kamer. Sledi natančnejši opis baze.

### 2.2.1 Pridobivanje slik

Baza SCface je bila ustvarjena predvsem z namenom testiranja nad slikami, ki so bile pridobljene pri pogojih, kot jih srečamo v resničnem svetu. V taki situaciji si je moč predstavljati scenarij, kjer je posameznika potrebno identificirati s primerjanjem visoko kvalitetne frontalne slike z nizko kvalitetno sliko iz nadzorne kamere. Za simulacijo opisanega scenarija baza SCface za vsakega posameznika vsebuje visoko kvalitetno frontalno sliko, tako v navadnem kakor v nočnem načinu, preostale slike pa so narejene s petimi manj kvalitetnimi kamerami, prav tako v obeh načinih. Edini vir svetlobe je bila svetloba iz stranskega okna.

Med pridobivanjem slik so se osebe morale sprehoditi po ravni črti, ki je vodila proti kameram, ki so se nahajale na višini 2,25 m, vmes pa se ustaviti na treh različnih razdaljah do kamer, natančneje pri 4,2 m, 2,6 m ter 1 m in počakati, da so jih kamere posnele. Opisan protokol je prikazan na sliki 2.2.

Po končanem pridobivanju slik so avtorji slike obrezali tako, da je na njih vidno le področje obraza, ki zakriva približno 80% celotne slike. Z namensko programsko opremo so tudi proizvedli dokument, ki za vsako sliko vsebuje koordinate oči, nosa in ust.



Slika 2.2: Razdalje, pri katerih so bile slikane osebe iz baze SCface [19].

## 2.2.2 Struktura baze in lastnosti slik

Bazo SCface so avtorji razdelili na pet glavnih sklopov in sicer na visoko kvalitetne frontalne slike v navadnem načinu, visoko kvalitetne slike v nočnem načinu, slike iz nadzornih kamer v navadnem načinu, slike iz nadzornih kamer v nočnem načinu ter slike s strani. Slednjih pri našem delu nismo uporabili, saj so namenjene predvsem testiranju 3D tehnik in jih zato tudi ne bomo več omenjali. Sledi podroben opis prvih štirih sklopov, primere pa lahko opazujemo na sliki 2.3.

### Visoko kvalitetne frontalne slike v navadnem načinu

Kot že rečeno vsebuje baza visoko kvalitetno frontalno sliko za vsakega izmed 130-ih posameznikov. Slike so velikosti  $1,600 \times 1,200$  pikslov in so takšne kvalitete, kakršno gre pričakovati pri slikah v policijskih kartotekah ali pri registraciji v varnostni sistem. Slike so označene po pravilu X.frontal, kjer X pomeni zaporedno številko osebe.

### Visoko kvalitetne frontalne slike v nočnem načinu

V tem sklopu so visoko kvalitetne frontalne slike vseh 130-ih oseb v nočnem načinu. Njihova ločljivost je  $426 \times 320$  pikslov, poimenovane pa so po pravilu X.cam8, kjer X pomeni zaporedno številko osebe.

### Slike iz nadzornih kamer v navadnem načinu

Vsakega izmed 130 posameznikov je posnelo pet kamer pri treh različnih razdaljah, kar pomeni, da ta sklop vsebuje 15 slik za vsakega posameznika. Kot je razvidno iz slike 2.3, se slike bistveno razlikujejo po kvaliteti in ločljivosti. Slike so označene po pravilu X.camY\_Z, kjer X pomeni zaporedno številko osebe, Y pomeni zaporedno številko kamere, Z pa nakazuje razdaljo, pri kateri je bila oseba slikana. Slike pri razdalji 1 so ločljivosti  $100 \times 75$  pikslov, pri razdalji 2  $144 \times 108$  pikslov, pri razdalji 3 pa  $224 \times 168$  pikslov. Za informacije o razdaljah glej sliko 2.2.

### Slike iz nadzornih kamer v nočnem načinu

Dve izmed kamer iz prejšnjega sklopa sta bili zmožni snemanja v nočnem načinu, kar pomeni, da za vsakega posameznika obstaja še šest infrardečih slik (dve kameri, tri razdalje). Kar zadeva ločljivost, so slike popolnoma enake slikam iz prejšnjega sklopa. Razlika je v poimenovanju, kjer se za poimenovanje kamer tu uporablja cam6 ter cam7.



Slika 2.3: Poln nabor slik iz baze SCface za eno osebo [19].

Če torej povzamemo, zakaj je baza obrazov SCface tako zanimiva za akademsko skupnost na področju prepoznavanja obrazov, so razlogi sledeči:

- Uporaba večih kamer z različno ločljivostjo in kvaliteto.

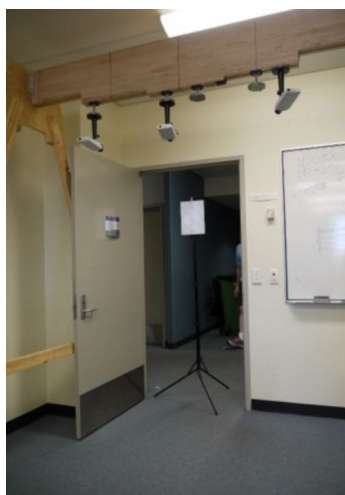
- Slike so bile narejene pri nenadzorovanih osvetlitvenih pogojih.
- Slike so bile narejene pri različnih razdaljah.
- Kamere so bile med pridobivanjem slik postavljene rahlo nad osebe, kar naredi prepoznavanje obrazov še bolj zahtevno. Poleg tega osebe niso gledale v fiksno točko.
- Baza vsebuje slike 130-ih oseb, kar pomeni, da je verjetnost pravilne prepoznave po naključju le 0.8%.

## 2.3 ChokePoint

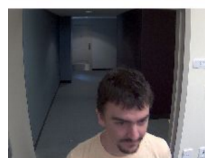
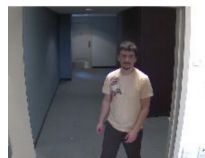
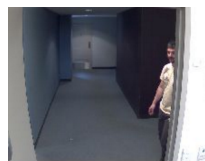
Baza obrazov ChokePoint je delo Wong et al. [64] in je namenjena testiranju video-to-video algoritmov za prepoznavanje obrazov. To pomeni, da so tako znani kot neznani obrazi predstavljeni z video sekvenco. Pri tem se baza razlikuje od baz SCface in FERET, ki sta namenjeni testiranju image-to-image algoritmov. Baza sestavlja 48 video sekvenc, ki vsebujejo kar 64,204 slik obrazov, vsekakor pa je na slikah le 30 različnih oseb, kar je velika pomanjkljivost baze.

### 2.3.1 Pridobivanje slik

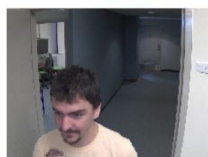
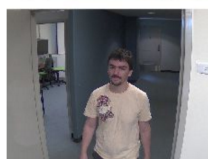
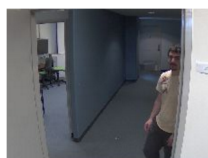
Pri en mesec trajajočem nastanku baze so bile uporabljene tri kamere, ki so bile postavljene nad oboki vrat, katera predstavljajo naravna ozka grla (ang. chokepoint) pri premikanju pešcev. Ko so se osebe na naraven način sprehodile skozi vrata, so bile posnete tako, da je vsaj ena izmed kamer obraz zajela frontalno oz. skoraj frontalno. Opisana postavitve kamer, naravna hoja oseb ter naravna osvetlitev so okoliščine, ki so zelo podobne tistim na javnih mestih, zaradi česar je baza ChokePoint zelo primerna za raziskave na področju prepoznavanja obrazov pri nenadzorovanih pogojih. Postavitve snemanja ter primere posnetkov omenjene postavitve lahko opazujemo na sliki 2.4.



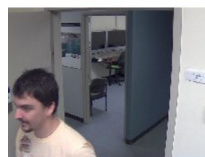
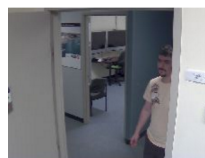
(a) Postavitev



(b) Kamera 1



(c) Kamera 2



(d) Kamera 3

Slika 2.4: Postavitev snemanja in primeri posnetkov v bazi ChokePoint [64].

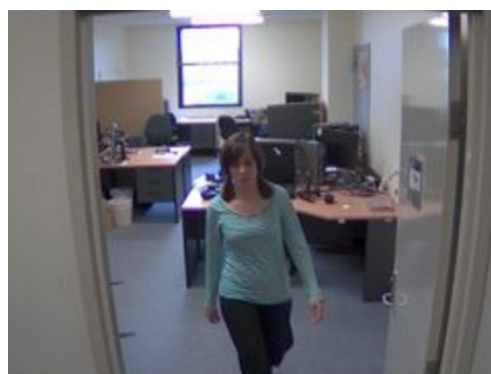
### 2.3.2 Struktura baze in lastnosti slik

Pri nastajanju baze se je 30 oseb večkrat sprehodilo skozi dvoje vrat v obe smeri, kjer so jih posnele tri kamere. Rezultat snemanja je 48 sekvenc, ki vsebujejo do 70 slik vsakega posameznika in so poimenovane po pravilu  $pWX\_sY\_cZ$ , kjer  $W$  pomeni številko vrat (1 ali 2),  $X$  sporoča smer sprehoda ( $E$  za prihod ter  $L$  za odhod),  $Y$  ponazarja številko sekvence,  $Z$  pa podaja

informacijo o kameri (1, 2 ali 3). Primer sekvence je npr. p1E\_s2\_c3, kar pomeni, da so osebe drugič (s2) prihajale skozi prva vrata (p1E), pri tem pa jih je posnela kamera številka tri (c3). Vse kamere so snemale pri hitrosti osveževanja 30 sličic na sekundo, sličice pa imajo ločljivost  $800 \times 600$  pikslov. Primere posnetkov iz različnih snemalnih postavitev lahko opazujemo na sliki 2.5.



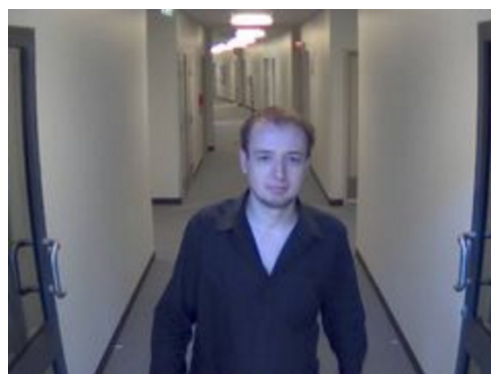
(a) P1E - Kamera 1



(b) P1L - Kamera 1



(c) P2E - Kamera 2

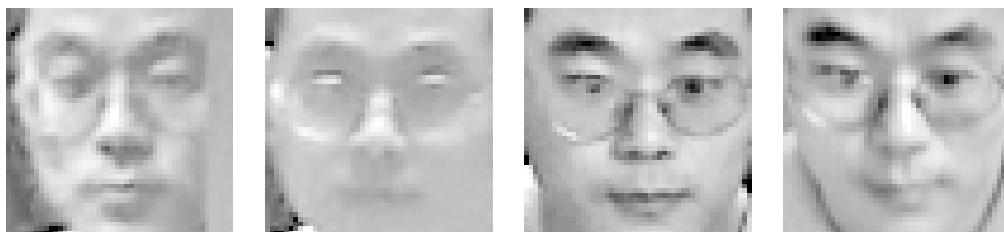


(d) P2L - Kamera 2

Slika 2.5: Primeri posnetkov iz različnih snemalnih postavitev v bazi ChokePoint [64].

Kot že omenjeno je baza ChokePoint namenjena testiranju video-to-video rešitvam za prepoznavanje obrazov, naš sistem pa deluje po principu image-to-image. To pomeni, da v koraku klasifikacije primerja eno sliko neznanega obraza s slikami znanih obrazov, pri čemer vsakemu znanemu obrazu pripada

le ena slika. Bazo smo za naš primer prilagodili tako, da smo za vsako osebo v dani sekvenci izračunali povprečen obraz. Rezultat povprečenja lahko opazujemo na sliki 2.6.



(a) Primeri štirih slik iste osebe v dani sekvenci, ki sicer vsebuje 58 slik.



(b) Povprečen obraz osebe v dani sekvenci

Slika 2.6: Da smo bazo ChokePoint prilagodili našemu sistemu, ki deluje po principu image-to-image, smo izračunali povprečen obraz posameznikov v vsaki sekvenci.



## Poglavje 3

# Metode prepoznavanja obrazov

V tem poglavju se bomo sprehodili skozi korake procesa prepoznavanja obrazov, ki so vidni na sliki 1.3 in pri tem za vsak korak opisali trenutno stanje v literaturi. Pri vsakem koraku bomo tudi podrobneje opisali metode, ki smo jih uporabili v naši rešitvi.

### 3.1 Detekcija obrazov

Za biometrične sisteme, ki kot vhod uporabljajo slike obrazov, pridobljene brez sodelovanja subjektov, je detekcija obraza ključnega pomena. Problem detekcije obrazov je zahteven saj mora sistem najti področje slike, ki vsebuje le obraz in to ob upoštevanju vseh variacij pri osvetlitvi, obraznih značilnostih, prekrivanju, itd. Nadalje mora delovati na obrazih, ki se pojavljajo na različnih razdaljah od kamere, v različnih pozah ter pri različnih rotacijah. Kljub temu je bil napredek na tem področju v zadnjem desetletju ogromen in mnogi sistemi so demonstrirali impresivno delovanje v realnem času [70].

Večina rešitev za detekcijo obrazov izvrši nalogo z ekstrakcijo določenih značilnosti iz množice učnih slik, ki vsebujejo obraze pri stalni oz. fiksni pozi [70]. Najpogosteje so to frontalne slike obrazov. Ko modul za detekcijo obrazov nato prejme sliko, na kateri mora zaznati obraz, to sliko preišče na

vseh lokacijah in povečavah ter na podlagi predhodno pridobljenih lastnosti zazna obraz. Po opisanem protokolu delujejo najnovejše rešitve, o katerih so poročali v delih Viola in Jones [62], Rowley et al. [46], Sung in Poggio [53], Schneiderman in Kanade [50] ter Yang et al. [71]. Na sliki 3.1 je ponazorjena uspešna detekcija obraza z metodo ERT [24].



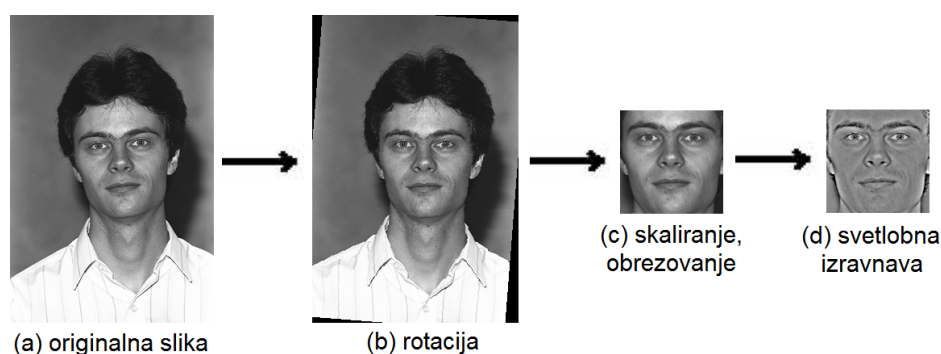
Slika 3.1: Uspešna detekcija obraza z metodo ERT [24].

### Uporabljena metoda

Pri našem delu detekciji obraza nismo posvečali posebne pozornosti saj je bila ta naloga za nas že rešena. Vse uporabljene baze obrazov so namreč vsebovale slike, ki prikazujejo le obraz osebe. Poleg tega so bile zraven priložene še datoteke, ki so vsebovale koordinate oči, nosu ter ust, kar je omogočilo učinkovito obrezovanje obrazov. Seveda je postopek ročnega določanja obraznih značilk zamuden ter je zanj potrebna človeška interakcija, zato bo potrebno sistem, ki je opisan v tej tezi, za uporabo na javnih mestih nadgraditi še z modulom za avtomatsko detekcijo obrazov.

## 3.2 Normalizacija

Ljudje imamo sposobnost prepoznati obraz ne glede na njegovo pozo, osvetlitev in izraz. Računalniki te zmožnosti ne posedujejo, zato jim je potrebno pri tem pomagati s predhodno obdelavo oz. normalizacijo slik. Pri tem procesu se na sliki minimizira vpliv lastnosti, ki niso del obraza ter s katerimi se oseba ne identificira. Pri omenjenih lastnostih imamo v mislih predvsem zgoraj naštetе pozo, osvetlitev ter izraz. Proces normalizacije je tradicionalno sestavljen iz rotacije, skaliranja, obrezovanja ter svetlobne izravnave slik [19] in ga lahko opazujemo na sliki 3.2, v zadnjih letih pa se v literaturi vse pogostejše pojavlja tudi frontalizacija [21, 18]. V naši rešitvi smo na bazi SCface rezultate izboljšali, če smo dodali še korak glajenja robov s konvolucijsko matriko. V tem razdelku bomo na hitro opisali trenutno stanje v literaturi na področju normalizacije slik obrazov ter metode, ki smo jih uporabili v naši rešitvi.



Slika 3.2: Proces normalizacije v naši rešitvi.

### 3.2.1 Rotacija

Korak rotacije je namenjen poravnavi obraza glede na izbrane značilke. Najpogostejše se rotacija izvaja glede na oči in sicer tako, da sta po končani rotaciji očesi na isti vodoravnici. Rotacija obraza je torej zelo trivialen proces, ki se ne razlikuje glede na različne metode normalizacije v literaturi in

sestoji iz “oblazinjenja” matrike slike z ničelnimi vrednostmi, translacije ter matričnega množenja [31]. Rezultat po rotaciji lahko opazujemo na sliki 3.2 pod točko (b).

### Uporabljen metoda

Pri naši rešitvi smo obraze rotirali okoli levega očesa, kar sestoji iz sledečih podoperacij:

- Obdajanje matrike  $I$  z ničlami za namen ohranitve originalnih dimenzij po rotaciji. Tukaj in v naslednjih podprocesih je  $I$  matrika, ki vsebuje sivinske vrednosti vseh pikslov na sliki.
- Translacije matrike  $I$  tako, da je levo oko v središču slike.
- Matričnega množenja, ki predstavlja dejansko rotacijo. Pri tem koraku je prvotno potrebno izračunati rotacijsko matriko, kar poteka po formuli 3.1 [31]:

$$M_{rotate} = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \quad (3.1)$$

kjer se kot rotiranja  $\alpha$ , če imata levo in desno oko koordinate  $(x_1, y_1)$  ter  $(x_2, y_2)$ , izračuna po enačbi 3.2 [31]:

$$\alpha = \text{atan}\left(\frac{y_2 - y_1}{x_2 - x_1}\right) \cdot \frac{180}{\pi} \quad (3.2)$$

Operacija rotacije se za vsak piksel  $I(x, y)$  na položaju  $(x, y)$  nato izvede z matričnim množenjem po pravilu 3.3 [31]:

$$I_r(x, y) - [x_c \ y_c] = [(x - x_c)(y - y_c)] \cdot M_{rotate} \quad (3.3)$$

kjer koordinati  $(x_c, y_c)$  označujeta središče slike (levo oko),  $I_r(x, y)$  pa predstavlja koordinate  $I(x, y)$  v novi, rotirani sliki.

- Translacije matrike tako, da je levo oko na prvotnih koordinatah.

### 3.2.2 Frontalizacija

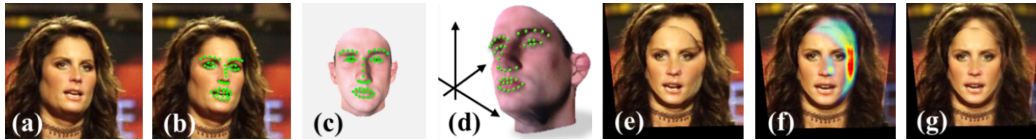
Eden največjih problemov, s katerimi se soočajo sistemi za prepoznavanje obrazov ob delovanju pri nenadzorovanih pogojih je položaj obraza. Zelo redko se namreč zgodi, da se na javnem mestu oseba zazre naravnost v kamero. Najpogostejše kamera osebo posname od zgoraj ter od strani. Da bi omenjene okoliščine čim manj vplivale na učinkovitost sistemov za prepoznavanje obrazov, so se v zadnjih letih pojavile rešitve za frontalizacijo obraza.

Frontalizacija je proces sintetiziranja frontalne slike obraza, zajetega pri nenadzorovanih pogojih [21]. V preteklosti so rešitve za frontalizacijo poskušale izračunati 3D model obraza za vsako obrazno sliko [54, 15, 20], a takšne metode imajo številne slabosti [21]. Kot prvo potrebujejo kvalitativne ter skoraj frontalne slike obrazov, čemur je v praksi zelo težko zadostiti. Nadalje so zelo občutljive na prekrivanja (npr. sončna očala) in zahtevajo natančno segmentacijo obraza od njegovega ozadja. Še ena slabost tiči v visoki odvisnosti od natančne lokalizacije obraznih značilk (oči, usta, nos), kar je pri nenadzorovanih pogojih velik problem [54]. Posledica uporabe takšnega pristopa na javnih mestih je, da za različne slike istega obraza proizvede zelo različne rezultate [21]. Nazadnje pa so omenjene metode še računsko zelo zahtevne in s tem neprimerne za uporabo pri realnem času.

Zgoraj opisane problematike so se lotili Hassner et al. [21] in prišli do na videz zelo enostavne rešitve. Njena osnova temelji na opustitvi računanja 3D modelov za vsako sliko posebej in uporabi le enega, vnaprej določenega 3D modela za preslikavo obraza. S tem postopkom se izgubi zelo malo intrinzičnih značilnosti, največjo prednost pa predstavlja natančna poravnost vseh obrazov kljub ekstremnim razlikam v pozi na prvotnih slikah. Ravno slednje je tisto, kar, glede na analizo avtorjev, najbolj pripomore k izboljšavi rezultatov pri prepoznavanju obrazov. Negativna posledica uporabe le enega 3D modela so možna popačenja pri obliki obraza, a so avtorji kljub temu izboljšali rezultate v literaturi. To je moč razložiti s predpostavko, da najbolj informativen del obraza (t.i. T-območje [35], ki pokriva širšo površino okoli oči ter nosu) ostaja pri postopku relativno nespremenjen.

## Uporabljena metoda

V naši rešitvi smo se poslužili metode, opisane v prejšnjem odstavku. Kot že omenjeno uporablja metoda le en vnaprej pripravljen 3D model za preslikavo obraza v frontalni pogled. Postopek preslikave lahko opazujemo na sliki 3.3, rezultate pa na sliki 3.4. Najboljše rezultate smo sicer dosegli brez frontalizacije. Zakaj je temu tako, bomo obrazložili v naslednjem poglavju pri analizi testov.



Slika 3.3: Proces frontalizacije pri rešitvi od Hassner et al. [21]: (a) Originalna slika; (b) Detekcija obraznih značilnk; (c) Isti detektor kakor v prejšnjem koraku je uporabljen za detekcijo značilnk na sliki obraza referenčnega 3D modela (d); (e) Iz 2D koordinat na originalni sliki in ustreznih 3D koordinat na modelu se izračuna projekcijska matrika, ki je nato uporabljena za projekcijo barv na referenčni model. (f) Slika prikazuje izračun območja obraza na frontalni sliki, ki ni bilo vidno na originalni sliki. Toplejše barve prikazujejo manj vidne piksele. Izgled obraza na tem območju je pridobljen z “izposojjo” barv iz druge strani obraza. (g) Končni rezultat frontalizacije.

### 3.2.3 Skaliranje

Pri skaliranju se spreminja število dimenzij matrike, ki opisuje sliko. Ker se v matlabu operacija izvede s preprostim ukazom `imresize(I, resize_ratio)`, kjer `I` predstavlja vhodno sliko, `resize_ratio` pa količnik spreminjanja, se na tem področju v širši pregled literature nismo spuščali. Rezultat operacije lahko opazujemo na sliki 3.2 pod točko (c).



(a) Slike pred frontalizacijo



(b) Slike po frontalizaciji

Slika 3.4: Primeri originalnih slik ter slik po frontalizaciji. Iz leve proti desni si sledijo slike iz baz FERET, SCface ter ChokePoint.

### Uporabljena metoda

V naši rešitvi smo slike skalirali tako, da je bila po končani operaciji pri vseh obrazih razdalja med očmi enaka. O konkretnih številkah pri razdalji bomo govorili v naslednjem poglavju pri analizi rezultatov.

### 3.2.4 Obrezovanje

Obrezovanje pri koraku normalizacije pomeni odstranjevanje področja slike, ki ne predstavlja obraza. Rezultat koraka je torej pravokotna slika, ki vsebuje le področje obraza kakor je vidno na sliki 3.2 pri koraku (d). Tako kot rotacija in spreminjane velikosti je tudi obrezovanje trivialna operacija in se ne razlikuje glede na različne metode normalizacije.

### Uporabljena metoda

Pri naši rešitvi smo si pomagali z datotekami, ki so bile priložene k bazam obrazov in so vsebovale koordinate oči, nosu ter ust. Obraz smo obrezali glede

na koordinati levega očesa in sicer tako, da je bilo med levim robom slike in levim očesom `left_margin` pikslov, med zgornjim robom slike in levim očesom pa `top_margin` pikslov. Naštete dimenzije smo izbrali na podlagi rešitev iz literature, kjer so teste izvajali na bazah FERET in SCface [32, 33, 34]. Pri tem smo pri bazah FERET in SCface za vrednosti `left_margin` in `top_margin` določili tako, da sta bila njuna deleža v primerjavi s širino in višino slike enaka 0.29 ter 0.27. Obrezovanje smo izvedli v matlabu s preprostim ukazom `imcrop(I,[x_min y_min width height])`, kjer `I` predstavlja vhodno sliko, `x_min` ter `y_min` sta koordinati zgornjega levega roba obrezovalnega pravokotnika, `width` in `height` pa predstavljata njegovo širino in višino.

### 3.2.5 Glajenje robov

Pri obdelavi slik se glajenje robov izvaja s konvolucijo med sliko ter ustrezno konvolucijsko matriko. Konvolucija je proces množenja vsakega piksla v sliki s svojimi sosednjimi piksli, ki so obteženi s konvolucijsko matriko. Če imamo na primer dve matriki velikosti  $3 \times 3$ , kjer prva predstavlja del slike, druga pa konvolucijsko matriko, se vrednost  $[2,2]$  elementa matrike po konvoluciji izračuna po pravilu 3.4:

$$\begin{aligned} & \left( \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \right) [2,2] \\ &= (i * 1) + (h * 2) + (g * 3) + (f * 4) + (e * 5) + \\ & \quad + (d * 6) + (c * 7) + (b * 8) + (a * 9). \end{aligned} \tag{3.4}$$

Najbolj pogost primer konvolucijske matrike, ki se uporablja za glajenje, je prikazan v razdelku 3.5:

$$\frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}. \tag{3.5}$$



## Uporabljena metoda

V naši rešitvi smo se glajenja poslužili samo pri bazi SCface, saj na bazah FERET in ChokePoint operacija ni izboljšala rezultatov. To je verjetno posledica tega, da so na slikah v bazi FERET robovi jasni ter nepopačeni, pri bazi ChokePoint pa se robovi zgladijo s procesom povprečenja slik, ki je razložen pri opisu baze v razdelku 2.3. Konvolucija se v Matlabu izvede preprosto z ukazom `conv2(A, B)`, kjer `A` predstavlja sliko, `B` pa konvolucijsko matriko, ki je bila v našem primeru kar vrstični vektor z le dvema dimenzijama in sicer  $\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$ .

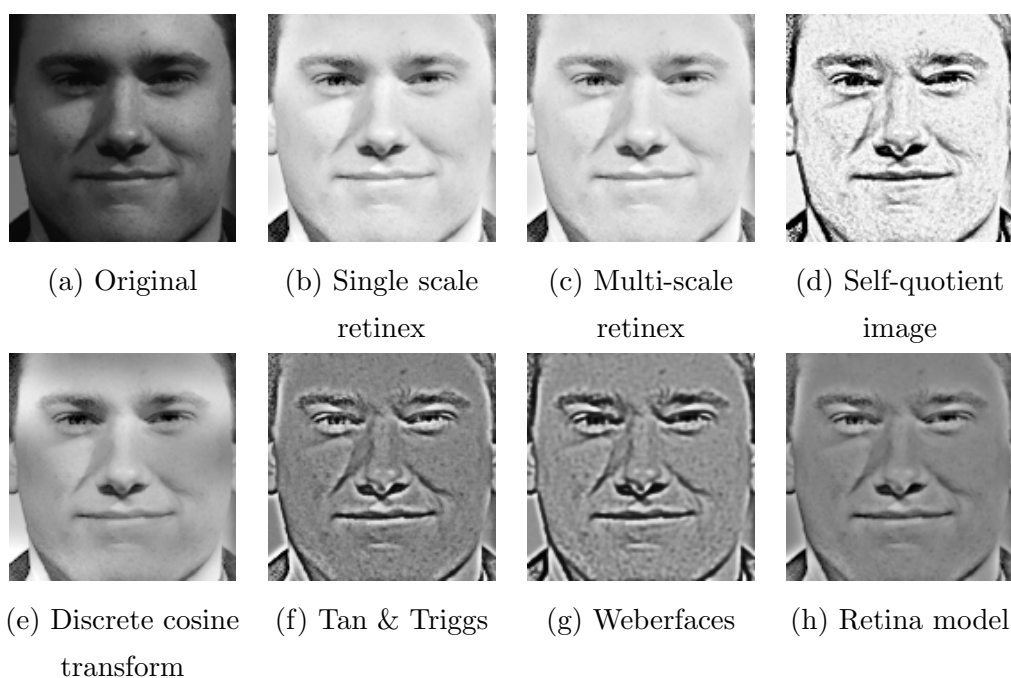
### 3.2.6 Svetlobna izravnava

Namen svetlobne izravnave je odstranjevanje vpliva osvetlitve obraza. Pri osvetlitvi obraza govorimo predvsem o intenzivnosti osvetlitve ter smeri, iz katere prihaja. Osvetlitev je pomemben dejavnik saj lahko poda zelo različne slike iste osebe in s tem bistveno vpliva na učinkovitost sistemov za prepoznavanje obrazov.

Čeprav so se v zadnjih letih v literaturi pojavile številne napredne metode za svetlobno izravnavo [30, 36], v ospredju ostajajo t.i. fotometrične metode, ki temeljijo na teoriji mrežnice (ang. Retinex theory). Ta pravi, da ima znana barva objektov v naravnem okolju povezavo z odbojnostjo, ki ostaja konstantna pri različnih osvetlitvenih pogojih [26]. Fotometrične rešitve so deležne velike pozornosti saj so računsko učinkovite ter ne potrebujejo posebnih učnih primerov. Največkrat uporabljene fotometrične rešitve v literaturi so [59, 5]: Retina model, metoda “Single scale retinex”, metoda “Multiscale retinex”, metoda “Self-quotient image”, Diskretna kosinusna transformacija, Tan & Triggs, metoda “Weberfaces”.

Poleg fotometričnih metod obstajajo tudi druge, ki so jih v delu Ochoa-Villegas et al. [30] poimenovali “osvetlitvene metode” (ang. relighting). Omenjene metode se učijo iz slik znanih obrazov, kjer naj bi bila osvetlitev nadzorovana, ter nato znane obraze osvetlijo tako, da čim boljše posnemajo

osvetlitev iz slik znanih obrazov. Metode so obetavne, a so slabše raziskane, računsko zahtevnejše, terjajo napredno opremo in potrebujejo množico učnih slik, s čimer niso primerne za rešitve, ki morajo delovati v realnem času. Poleg tega pa se danes v sistemih prepoznavanja obrazov uporablja napredne rešitve v koraku ekstrakcije značilk, ki pa so zelo odporne na variacije pri osvetlitvi in jih bomo opisali v razdelku 3.3.



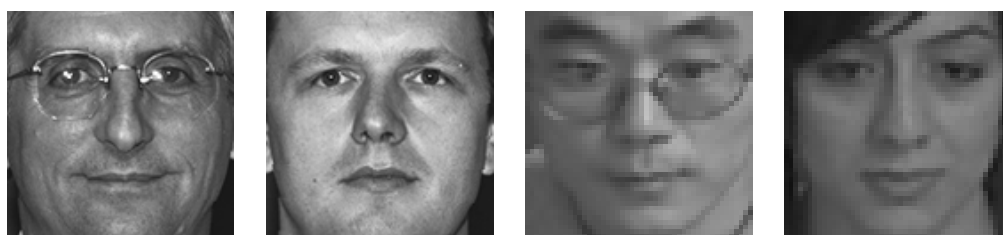
Slika 3.5: Primeri rezultatov fotometričnih metod.

### Uporabljena metoda

Pri naši rešitvi smo zaradi njihove računske učinkovitosti razmišljali le o fotometričnih metodah. Pri testiranju smo pregledali knjižnico INFace [58], ki vsebuje vse poprej naštet metode in še nekatere druge. Njihove rezultate lahko opazujemo na sliki 3.5. Najboljše rezultate sta podali metodi Retina model pri bazah FERET in ChokePoint ter navadna histogramna izravnava pri bazi SCface.

## Retina model

Metoda Retina model temelji na teoriji mrežnice in sestoji iz apliciranja dveh adaptivnih nelinearnih funkcij ter razlike Gaussovega filtra [63]. Postopek s tem posnema delovanje človeške mrežnice, ki ji prepoznavanje barv ob različnih osvetlitvenih pogojih omogočata dva sloja in sicer fotoreceptorska plast ter zunanja plexiformna plast [63]. Primere slik pred in po aplikaciji metode Single scale retinex lahko opazujemo na sliki 3.6.



(a) Originalne slike

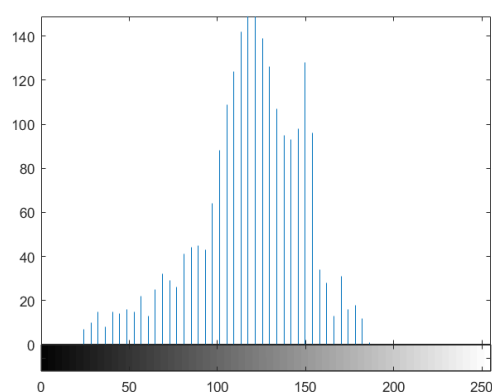


(b) Slike po svetlobni izravnavi z metodo Retina model

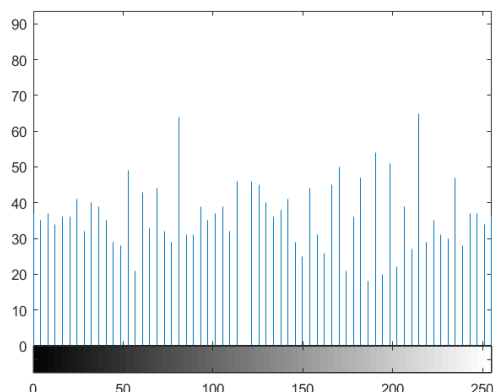
Slika 3.6: Primeri originalnih slik ter slik po aplikaciji metode Retina model. Na levi strani so slike iz baze FERET, na desni pa slike iz baze ChokePoint.

## Histogramska izravnava

Histogramska izravnava je tehnika, ki se uporablja za namen povečave kontrasta [44]. To stori tako, da učinkovito razporedi najbolj pogoste intenzivnosti na celoten histogram, kar lahko opazujemo na sliki 3.7. S tem se območjem z nižjim kontrastom le ta poveča. Primere slik pred in po histogramski izravnavi lahko opazujemo na sliki 3.8.



(a) Histogram originalne slike



(b) Histogram slike po histogramski izravnavi

Slika 3.7: Histogram iste slike pred in po histogramski izravnavi. Na horizontalni osi so navedene intenzivnosti (od 0 do 255), na vertikalni pa število njihovih pojavitev.



(a) Originalne slike



(b) Slike po histogramski izravnavi

Slika 3.8: Primeri originalnih slik ter slik po histogramski izravnavi. Vse slike so iz baze SCface.

Kot že omenjeno, je bila histogramska izravnava uporabljena le na bazi SCface. Na bazi smo preizkusili vse fotometrične metode iz knjižnice INFace, a je ta enostavna tehnika podala najboljše rezultate. Razlog tiči v tem, da so slike v bazi SCface nizke kvalitete, zamegljene in vsebujejo malo drobnih podrobnosti.

### 3.3 Extrakcija značilk

Ekstrakcija značilk je najbolj pomemben korak v procesu prepoznavanja obrazov [31]. Tukaj sistem po določenem postopku spremeni slike v obliko, v kakršni se jih nato med seboj primerja v koraku klasifikacije. Najpogostejše je ta oblika vektor. Glede na način ekstrakcije značilk je možno metode kategorizirati v dve glavni skupini [72]:

- **Globalne oz. celostne ali holistične metode:** te metode v koraku ekstrakcije značilk proizvedejo vektor, ki vsebuje informacije o vzorcih in lastnostih celotnega obraza. Najbolj pogosto se v literaturi omenjata metodi Eigenfaces [61] in Fisherfaces [8].
- **Lokalne metode:** Za razliko od globalnih metod, ki se osredotočajo na celoten obraz, te metode temeljijo na segmentaciji obraza v različne lokalne značilke kot so oči, usta, nos itd. Vsaka slika je nato predstavljena s t.i. vektorjem značilk, pridobljenim z določenim ekstrakcijskim algoritmom, ki iz omenjenih lokalnih značilk izlušči najbolj informativne lastnosti. V literaturi so najbolj popularne metode LBP - Lokalni binarni vzorci (ang. *Local Binary Patterns*) [2], Gaborjevi filtri [17] in LPQ - Kvantizacija lokalnih faz (ang. *Local Phase Quantization*) [4].

V primerjavi z globalnimi imajo lokalne metode ključno prednost [31]: so veliko bolj učinkovite pri delovanju ob nenadzorovanih pogojih. Po poročanju Yan et al. [68] so lokalne metode 60% bolj uspešne pri spopadanju s pozami pri rotaciji za  $\pm 40^\circ$ . Nadalje, kot so potrdili Sinha et. al [52], se izgled nekaterih lokalnih značilk ne spreminja s pozo, obraznim izrazom in svetlobno

usmeritvijo. Dodatno so Sadr et al. [47] ugotovili, da je moč z le eno lokalno značilko, kot so npr. obrvi ali oči, pravilno klasificirati obraze slavnih oseb.

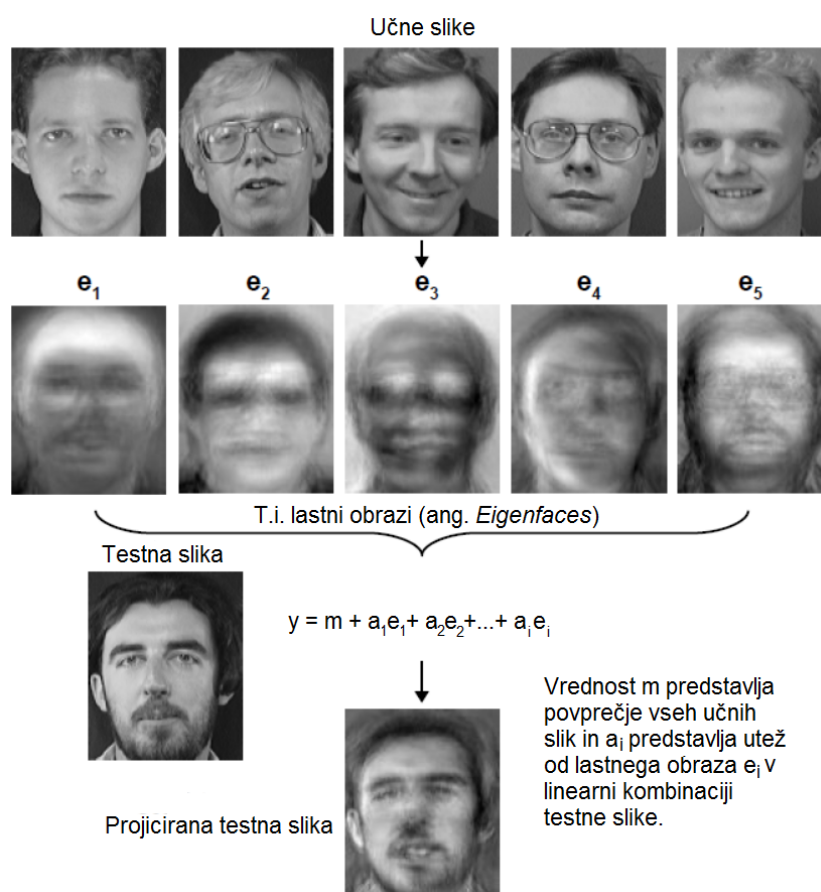
Zaradi opisanih prednosti lokalnih metod smo se v tem delu posvetili prav njim. Vseeno smo rešitev testirali tudi z eno globalno metodo, natančneje Eigenfaces, da smo lahko opazovali razlike v rezultatih pri uporabi različnih normalizacijskih tehnik. Sledi podrobnejši opis uporabljenih metod.

### 3.3.1 Eigenfaces

Eigenfaces [61] je najbolj znana holistična metoda za ekstrakcijo značilk. Metoda je osnovana na predpostavki, da je vsako sliko obraza iz dane množice možno predstaviti kot linearno kombinacijo ortogonalnih lastnih vektorjev, izračunanih z Metodo glavnih komponent (PCA) na omenjeni množici [25].

Na začetku je vsaka slika predstavljena z matriko sivinskih vrednosti velikosti  $M \times N$ , kjer je  $M$  število vrstic in  $N$  število stolpcev. Če je slika ločljivosti  $48 \times 48$ , imata  $M$  in  $N$  vrednost 48. Prvi korak pri metodi Eigenfaces je, da se omenjena matrika preoblikuje v stolpični vektor dolžine  $M * N$ . V naslednjem koraku se vektorju vsake slike odšteje povprečni vektor celotne množice slik. S tem se od vektorjev odstrani skupne informacije, ki ne pripomorejo k učinkovitejši klasifikaciji obrazov. Nato se iz množice vektorjev z metodo PCA izračuna lastne vektorje in se z njimi sestavi projekcijsko matriko tako, da so lastni vektorji urejeni po velikosti od leve proti desni. S to matriko se vsak originalni vektor projicira v nov vektorski prostor, kjer je vsaka slika predstavljena z linearno kombinacijo pridobljenih lastnih vektorjev. Metoda se imenuje Eigenfaces, ker je lastne vektorje moč rekonstruirati in jih prikazati kot obraze, kakor je to vidno na sliki 3.9.

Metoda podaja dobre rezultate [60] pri množicah visoko kvalitetnih slik, ki so bile pridobljene pri nadzorovanih pogojih ter ne izkazujejo sprememb, ki so posledica staranja ali zakrivljanja. V nasprotnem primeru njihova učinkovitost drastično pade. V delu Grgić et al. [19] so npr. dosegli le 4.7% uspešnost pri uporabi opisane metode na bazi SCface.



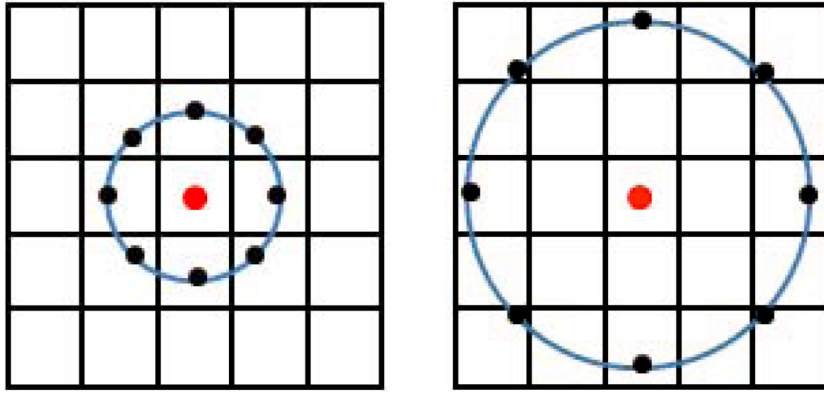
Slika 3.9: Postopek metode Eigenfaces [61].

### 3.3.2 LBP in ELBP

LBP je poleg Gaborjevih filtrov največkrat omenjena lokalna metoda v literaturi, in to z dobrim razlogom: je preprosta, računsko ne-zahtevna ter odporna do sprememb pri osvetlitvi [32]. Metoda prvotno ni bila razvita za uporabo pri prepoznavanju obrazov pač pa za prepoznavanje tekstur [37]. Kot prvi so za prepoznavanje obrazov metodo uporabili Ahonen et. al [3] in dosegli zelo obetavne rezultate. Od takrat se je metoda LBP na področju prepoznavanja obrazov razširila na uporabo pri detekciji obrazov, prepoznavanju obraznih izrazov, določanju starosti, določanju spola [32, 3] itd.

### Postopek ekstrakcije značilnk z metodo LBP [32]:

Za vsak piksel na lokaciji  $(x_c, y_c)$  se LBP vrednost izračuna s primerjavo njegove sivinske vrednosti  $g_c$  s sivinskimi vrednostmi  $P$  okoliških pikslov, ki se nahajajo na krožnici s polmerom  $R$ . Če za ekstrakcijo značilnk uporabimo metodo LBP s primerjavo osmih okoliških pikslov ( $P = 8$ ), ki ležijo na krožnici s polmerom 1 ( $R = 1$ ), to zapišemo z oznako  $LBP^{8,1}$ . Vizualizacijo primera lahko opazujemo na sliki 3.10.



Slika 3.10: Vzorca  $LBP^{8,1}$  in  $LBP^{8,2}$ .

LBP vrednost piksla na lokaciji  $(x_c, y_c)$  se izračuna po enačbi 3.6:

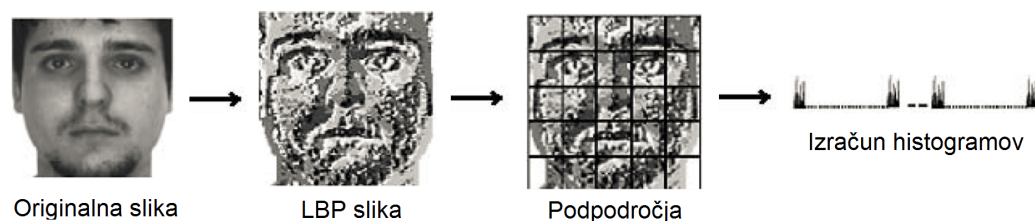
$$LBP^{P,R}(x_c, y_c) = \sum_{i=1}^P s(g_i^{P,R} - g_c) 2^{i-1} \quad (3.6)$$

kjer je  $s(x)$  definiran po pravilu 3.7:

$$s(x) = \begin{cases} 1 & \text{če } x \geq 0 \\ 0 & \text{če } x < 0. \end{cases} \quad (3.7)$$



V prvem koraku ekstrakcije značilke se torej za vsako sliko z zgoraj opisano enačbo izračuna njena LBP predstavitev, ki jo bomo poimenovali slika LBP. Nato se slika LBP razdeli na manjša, ne-prekrivajoča se podpodročja. V naslednjem koraku se izračunajo histogrami omenjenih podpodročij, ki se zapišejo kot vektorji, le ti pa se na koncu združijo v en sam vektor, ki predstavlja končni rezultat. Proces lahko opazujemo na sliki 3.11.

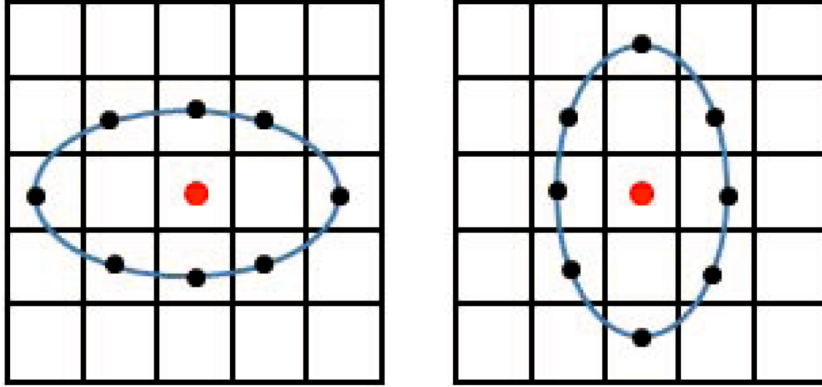


Slika 3.11: Koraki v procesu računanja LBP vektorja.

V delu Nguyen et al. [32] so ugotavljali, da imajo najbolj informativne obrazne značilke, predvsem oči in usta, obliko elipse. Na podlagi tega so poizkusili piksele primerjati ne po krožnici, pač pa po elipsi. Dognali so, da se rezultati izboljšajo, če se piksli primerjajo po horizontalno ležeči elipsi. Rezultate so še dodatno izboljšali, ko so predstavitev, izračunani po vodoravni elipsi dodali še predstavitev, izračunano po vertikalni elipsi. V procesu so iznašli nadgradnjo metode LBP, ki so jo poimenovali Eliptični lokalni binarni vzorci (ang. *Eliptical Local Binary Patterns* - ELBP).

### Postopek ekstrakcije značilke z metodo ELBP [32]:

Postopek je zelo podoben kot pri metodi LBP. Za vsak piksel na lokaciji  $(x_c, y_c)$  se ELBP vrednost izračuna s primerjavo njegove sivinske vrednosti  $g_c$  s sivinskimi vrednostmi  $P$  okoliških pikslov, ki se nahajajo na elipsi z osema  $R_1$  in  $R_2$ . Če za ekstrakcijo značilke uporabimo metodo ELBP s primerjavo osmih okoliških pikslov ( $P = 8$ ), ki ležijo na elipsi z osema 1 in 2 ( $R_1 = 1$ ,  $R_2 = 2$ ), to zapišemo z oznako  $ELBP^{8,1,2}$ . Vizualizacijo primera lahko opazujemo na sliki 3.12.



Slika 3.12: Vzorca  $ELBP^{8,2,1}$  in  $ELBP^{8,1,2}$ . Ko je  $R1 = R2$  je ELBP isti kot LBP; ko je  $R1 > R2$ , se primerjajo piksli na horizontalni elipsi; ko je  $R1 < R2$ , se primerjajo piksli na vertikalni elipsi.

ELBP vrednost piksla na lokaciji  $(x_c, y_c)$  se izračuna po enačbi 3.8:

$$ELBP^{P,R1,R2}(x_c, y_c) = \sum_{i=1}^P s(g_i^{P,R1,R2} - g_c) 2^{i-1} \quad (3.8)$$

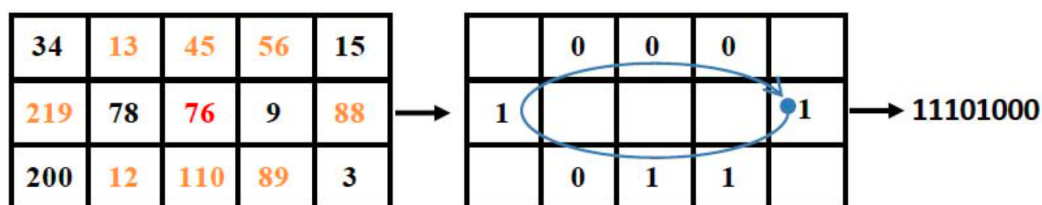
kjer je definicija  $s(x)$  enaka kot pri LBP. Postopek izračuna ELBP vrednosti za en piksel je prikazan na sliki 3.13. Koordinate  $i$ -tega piksla, ki leži na elipsi se izračuna po enačbah 3.9, 3.10 in 3.11:

$$angle\_step = 2 * \pi / P \quad (3.9)$$

$$x_i = x_c + R1 * \cos((i - 1) * angle\_step) \quad (3.10)$$

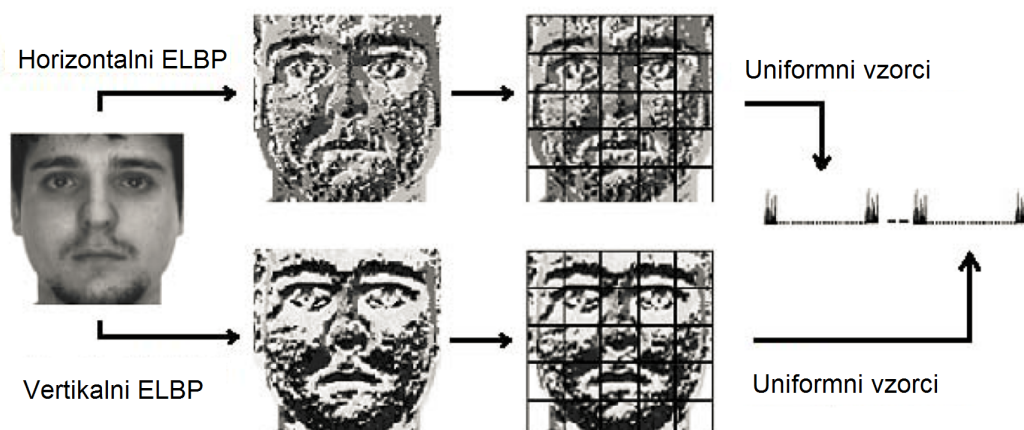
$$y_i = y_c + R2 * \sin((i - 1) * angle\_step) \quad (3.11)$$

Kot smo že dejali, se rezultati izboljšajo, če pri metodi izračunamo tako horizontalni kot vertikalni ELBP vektor, ki ju nato združimo v en  $ELBP(h+v)$  vektor, kakor so ga poimenovali v delu Nguyen et al. [32]. Tudi postopek računanja  $ELBP(h+v)$  vektorja je zelo podoben temu pri metodi LBP. Razlika je le v tem, da se postopek računanja ELBP slike opravi dvakrat; enkrat z uporabo horizontalne, drugič pa z uporabo vertikalne elipse. Za tem



Slika 3.13: Izračun ELBP vrednosti za en piksel pri uporabi deskriptorja  $ELBP^{8,2,1}$ .

se sliki ponovno razdeli na manjša podpodročja ter izračuna njihove histograme. Vektorji, ki opisujejo histograme obeh slik se nato združijo v en sam vektor, ki predstavlja končni rezultat. Proces lahko opazujemo na sliki 3.14. Ker se v rezultatu upoštevata dve predstavitvi (horizontalna in vertikalna), je vektor dvakrat daljši, kot pri LBP.



Slika 3.14: Koraki v procesu računanja  $ELBP(h+v)$  vektorja [32].

Omeniti je potrebno še, da se tako pri LBP kot pri ELBP histogrami izračunajo z upoštevanjem le uniformnih vzorcev [3]. Binarna vrednost je uniformni vzorec, če ima največ dva bitna prehoda iz 0 v 1 ali obratno. Z uporabo uniformnih vzorcev se dolžina histograma zmanjša iz 256 (število vseh sivinskih intenzivnosti) na 59 (število vseh uniformnih vzorcev pri dolžini osmih bitov). Če torej po koraku računanja ELBP slike le to razdelimo na  $W * H$

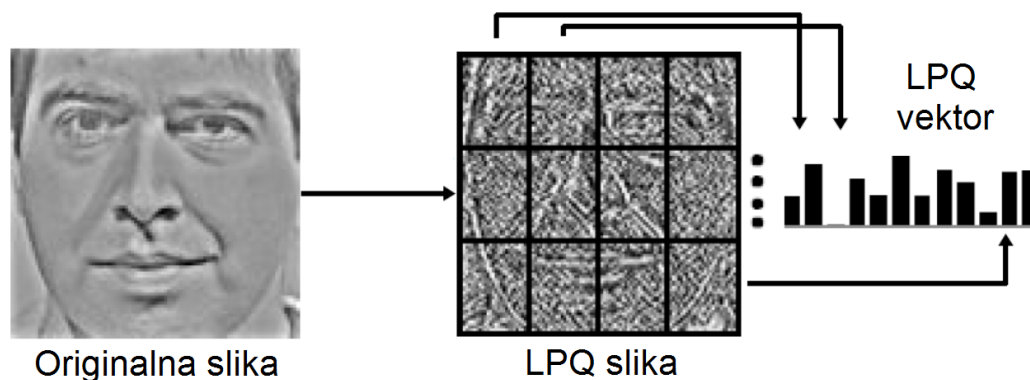
podpodročij, je dolžina končnega vektorja enaka  $W * H * 59$ . Seveda, če računamo ELBP(h+v) vektor, je dolžina dvakrat daljša. O vrednostih vseh spremenljivk, omenjenih pri opisu metod LBP in ELBP, bomo poročali pri analiziranju rezultatov.

### 3.3.3 LPQ

Kvantizacija fazne razlike (ang. *Local Phase Quantization* - LPQ) je metoda, ki je bila leta 2008 razvita za namen opisovanja tekstur [38], še istega leta pa so jo prvič uporabili tudi na področju prepoznavanja obrazov [4]. Metoda je osnovana na kvantizaciji fazne razlike Fourierjeve transformacije v lokalnih območjih. Omenjena lastnost signala je bila izbrana, ker je pri določenih pogojih invariantna na zamegljenost slike [4].

#### Postopek ekstrakcije značilnk z metodo LPQ [38]

Na osnovi invariantnosti fazne razlike na zamegljenost se LPQ predstavitev slike izračuna v dveh korakih: z aplikacijo kratkoročne Fourierjeve transformacije na lokalnih območjih slike ter s kvantizacijo pridobljenih koeficientov. Zatam se na lokalnih območjih pridobljene LPQ slike izračunajo še histogrami, ki se zapišejo v en sam vektor. Shemo procesa lahko opazujemo na sliki 3.15.



Slika 3.15: Koraki v procesu računanja LPQ vektorja [31].

V prvem koraku se torej aplicira kratkotrajna Fourierjeva transformacija na lokalnih območjih velikosti  $M \times M$  za vsak piksel  $x$ . To se stori z uporabo štirih 2-D frekvenc  $u_1 = [a, 0]^T$ ,  $u_2 = [0, a]^T$ ,  $u_3 = [a, a]^T$ ,  $u_4 = [a, -a]^T$ , kjer je  $a$  majhno število, ki zadovolji pogoj invariantnosti na zamegljenost [38]. Glede na priporočila iz dela Nguyen et. al [33] smo število  $a$  določili po pravilu 3.12:

$$a = \frac{8}{M-1}. \quad (3.12)$$

Z opisanim postopkom se za vsak piksel proizvede vektor, ki prikazan v razdelku 3.13:

$$F_x = [F(u_1, x), F(u_2, x), F(u_3, x), F(u_4, x)]. \quad (3.13)$$

Fazna razlika Fourierjevih koeficientov se pridobi z opazovanjem realnih in imaginarnih delov vseh komponent v  $F_x$ . To se opravi s preprosto kvantizacijo po pravilu 3.14:

$$q_j(x) = \begin{cases} 1 & \text{če } g_j(x) \geq 0 \\ 0 & \text{drugače} \end{cases}, \quad (3.14)$$

kjer je  $g_j(x)$   $j$ -ta komponenta vektorja  $G_x = [Re\{F_x\}, Im\{F_x\}]$ . Pridobljenih osem binarnih koeficientov se nato zapiše v desetiško obliko z binarnim kodiranjem po enačbi 3.15

$$f_{LPQ}(x) = \sum_{j=1}^8 q_j(x) 2^{j-1}. \quad (3.15)$$

V tem koraku dobimo t.i. LPQ sliko, ki jo namesto sivinskih intenzivnosti sestavljajo LPQ vrednosti. Če so parametri izbrani pravilno, je LPQ slika zelo odporna na zameglitev.

Preostanek metode je zelo podoben temu iz poprej opisanih metod LBP in ELBP. LPQ slika se razdeli na  $W \times H$  podpodročij, katerim se nato izračunajo histogrami. Vektorji, ki opisujejo histograme se zatem združijo v en sam vektor, ki predstavlja končni rezultat. Metoda LPQ se od metod LBP in ELBP

razlikuje po tem, da se pri računanju histogramov ne uporabijo uniformni vzorci. Posledično je LPQ vektor štirikrat daljši od LBP vektorja oz. dvakrat daljši od ELBP(h+v) vektorja. O konstantah, ki so bile omenjene pri opisu metode, bomo govorili pri analizi rezultatov.

### 3.3.4 LPOG

Metoda Lokalni vzorci gradientov (ang. *Local Patterns of Gradients* - LPOG), ki so jo leta 2015 razvili Nguyen et al. [34], združuje prednosti metod LPQ in ELBP ter je trenutno ena izmed vodilnih metod za ekstrakcijo značilnk v literaturi, sodeč po rezultatih na podatkovnih bazah SCface [19], AR [28] in FERET [43].

#### Postopek ekstrakcije značilnk z metodo LPOG [34]:

Iz vhodne slike se prvotno izračunata dve usmerjeni gradientni sliki, na katere se nato paroma aplicirata deskriptorja LPQ in BELBP (nadgradnja metode ELBP) za generacijo lokalnih vzorčnih slik. Zatem se izračunajo histogrami lokalnih območij vseh omenjenih slik, ki se na koncu združijo v en sam LPOG vektor. Kakor je razvidno iz sheme 3.16 prihajajo vsi koraki neposredno iz že opisanih metod LPQ in ELBP, a z dvema izjemama: izračun dveh usmerjenih gradientnih slik v prvem koraku ter nadgradnja metode ELBP.

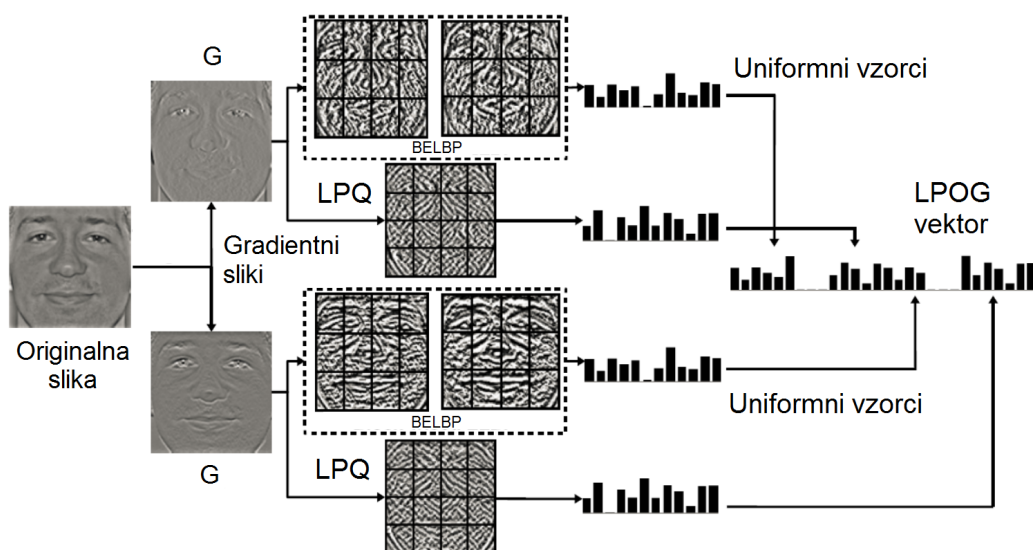
Iz vhodne slike  $I$  se torej v prvem koraku izračunata dve gradientni sliki  $G_x$  in  $G_y$ , po ena za vsako os  $x$  in  $y$ . Izračun poteka po formulah 3.16 in 3.17:

$$G_x = -\frac{1}{2} \cdot I(x-1, y) + 0 \cdot I(x, y) + \frac{1}{2} \cdot I(x+1, y) \quad (3.16)$$

$$G_y = -\frac{1}{2} \cdot I(x, y-1) + 0 \cdot I(x, y) + \frac{1}{2} \cdot I(x, y+1) \quad (3.17)$$

Gradientne slike vsebujejo okrepljene informacije o robovih, kar je zelo pomembno pri izgradnji predstavitve obraza. Nadalje doprinesejo k dodatni odpornosti deskriptorja na osvetlitev [34].

Na pridobljenih gradientnih slikah se nato aplicirata deskriptorja LPQ in BELBP, opise katerih najdemo v razdelkih 3.3.3 in 3.3.2, a je pri slednjem



Slika 3.16: Koraki v procesu računanja LPOG vektorja [34].

potrebna še dodatna razlaga. V razdelku 3.3.2 namreč najdemo opis metode ELBP, tukaj pa govorimo o metodi BELBP oz. bločni ELBP (ang. *Block-Wised ELBP*). Razlika tiči v tem, da se horizontalna in vertikalna predstavitev ELBP ne računata na vhodni sliki (oz. v primeru LPOG na gradientnih slikah  $G_x$  in  $G_y$ ) temveč na predhodno izračunanih akumuliranih slikah, ki se pridobita po enačbi 3.18:

$$AI_{v,h}^{BP}(x,y) = \frac{1}{N} \sum_{(x,y) \in BP} I(x,y) \quad (3.18)$$

kjer sta  $v$  in  $h$  notaciji za vertikalno in horizontalno usmeritev,  $N$  je število pikslov v vzorcu BP in  $I(x,y)$  je vrednost intenzivnosti na lokaciji  $(x,y)$ . Proizvedeta se torej dve akumulirani sliki, ena po x osi ter ena po y osi, na katerih se nato upoštevajoč usmeritev ustrezno aplicirata deskriptorja horizontalni ter vertikalni ELBP.

Preostanek metode LPOG je, kot rečeno, identičen temu pri metodah LPQ in ELBP. Pridobljene LPQ in BELBP predstavitve slik se razdelijo na  $W * H$  podpodročij iz katerih se nato izračunajo histogrami. Pri tem se na

BELBP slikah uporabijo uniformni vzorci. Vektorji, ki opisujejo histograme, se na koncu združijo v en sam LPOG vektor. Ker pridobljen vektor vsebuje LPQ in BELBP deskriptorje dveh gradientnih slik, je trikrat daljši od LPQ vektorja in šestkrat daljši od ELBP oz. BELBP vektorja. O neznanih konstantah, ki so bile omenjene v tem opisu, bomo govorili pri analizi rezultatov.

### 3.3.5 Povzetek metod za ekstrakcijo značilnk

V tem poglavju smo opisali uporabljene metode za ekstrakcijo značilnk, natančneje LBP [3], ELBP [32], LPQ [4] in LPOG [34]. Omenjene metode smo izbrali zaradi rezultatov, ki so jih dosegale na bazi obrazov SCface. Metodi LBP in ELBP sta robustni na osvetlitev, metoda LPQ pa podaja dobre rezultate pri prepoznavanju zamegljenih slik. Metoda LPOG združuje prednosti metod LPQ in ELBP, ki jih dopolni z uporabo gradientnih slik, s čimer se še poveča odpornost na osvetlitev, pripomore pa tudi k natančnejšemu opisu obraza zaradi izboljšane informacije o robovih. Empirično se metoda LPOG izkaže tudi pri prepoznavanju obrazov pri nizki ločljivosti slik [34]. Dodatno se z razdeljevanjem vzorčnih predstavitev slik na podpodročja poveča odpornost deskriptorjev na pozo, zakrivanje, izraz in staranje [34].

## 3.4 Zmanjšanje dimenzionalnosti

Ekstrakcija značilnk je najpomembnejši korak v postopku prepoznavanja obrazov [31]. V tem koraku se predstavitev obraza pretvori v vektor značilnk, ki se v fazi klasifikacije uporabi za primerjavo z drugimi obrazi. Na podlagi klasifikacije se obrazu na koncu podeli identiteta. A takšen sistem je zaradi visoke dimenzionalnosti vektorjev zelo počasen. Za namen odstranitve omenjene slabosti se uporabi učni algoritem, s čimer se povečata tako hitrost, kot tudi učinkovitost sistema, saj so podatki v novem vektorskem podprostoru bolj prepoznavni. V literaturi se na tem področju najbolj pogosto omenjata metodi Analiza linearne diskriminante (ang. *Linear Discriminant Analysis* -



LDA) [56] ter Metoda glavnih komponent (ang. *Principle Component Analysis* - PCA) [60]. Zaradi enostavnejše implementacije, večje hitrosti [29, 11] in boljših rezultatov v literaturi [32, 33, 34, 29] smo se pri našem delu odločili za slednjo.

## PCA

Metoda PCA je statistični postopek, ki z uporabo ortogonalne transformacije pretvori množico vrednosti na nov koordinatni sistem, čigar prva koordinata, imenovana prva glavna komponenta, izkazuje največjo varianco podatkov, druga komponenta drugo največjo varianco itd. Če ohranimo le  $n$  glavnih komponent, se zmanjša dimenzionalnost projiciranih podatkov ob minimalni izgubi variacije. Glede deleža variacije, ki ga je za učinkovito delovanje sistema potrebno ohraniti, najdemo v literaturi različna mnenja [19, 67, 41], a je vselej višji od 95%.

Proces zmanjšanja dimenzionalnosti z metodo PCA je sestavljen iz dveh korakov in sicer iz učenja in projiciranja. Rezultat prvega je projekcijska matrika, ki jo sestavljajo lastni vektorji, izračunani iz matrike vektorjev značilnih slik. Pridobljeno projekcijsko matriko se nato uporabi za projiciranje vseh vektorjev značilnih znanih in neznanih obrazov v nov vektorski podprostor, zgrajen iz učnih slik.

Projekcijsko matriko je moč izračunati na dva načina [31] in sicer z metodo razgradnje lastnih vrednosti (ang. *Eigenvalue Decomposition* - EVD) in z metodo razgradnje singularnih vrednosti (ang. *Singular Value Decomposition* - SVD). Po ugotovitvah Nguyen et al. [31] je slednja metoda računsko zanesljivejša in hitrejša, saj za razliko od metode EVD ni potreben izračun kovariančne matrike ter hranjenje matrike originalnih podatkov za izračun lastnih vektorjev. Posledično smo se pri naši rešitvi odločili za metodo SVD, katere opis sledi.

Če imamo  $K$  učnih slik, ki so predstavljene s stolpičnimi vektorji značilk  $x_i$  ( $i = 1..K$ ) in velja  $A = [\Phi_1, \Phi_2.. \Phi_K]$ , kjer je:

$$\Phi_i = x_i - \bar{x}, (i = 1..K), \bar{x} = \frac{1}{K} \sum_{i=1}^K x_i, \quad (3.19)$$

potem se projekcijsko matriko izračuna preprosto z razgradnjo singularnih vrednosti po pravilu 3.20:

$$[U, \Sigma, V] = svd(A), \quad (3.20)$$

kjer velja  $A = U\Sigma V$ . Pri tem so v  $U$  vsebovani lastni vektorji od  $A$  (po en lastni vektor na stolpec, razvrščen po velikosti od leve proti desni),  $\Sigma$  pa je diagonalna matrika, ki vsebuje singularne vrednosti, ki so v našem primeru enake kvadratnim korenem ustreznih lastnih vrednosti  $\Lambda$  (prav tako razvrščene po velikosti) oz. velja  $\Sigma^2 = \Lambda$ . Matrika  $V$  je enaka transpoziciji matrike  $U$  ter nas ne zanima. Že v tem koraku bi lahko uporabili  $U$  kot projekcijsko matriko, a bo nov vektorski podprostor bolj diskriminanten, če projekcijsko matriko predhodno pobelimo. To pomeni, da lastne vektorje pomnožimo z inverzom kvadratnih korenov pripadajočih lastnih vrednosti po pravilu 3.21:

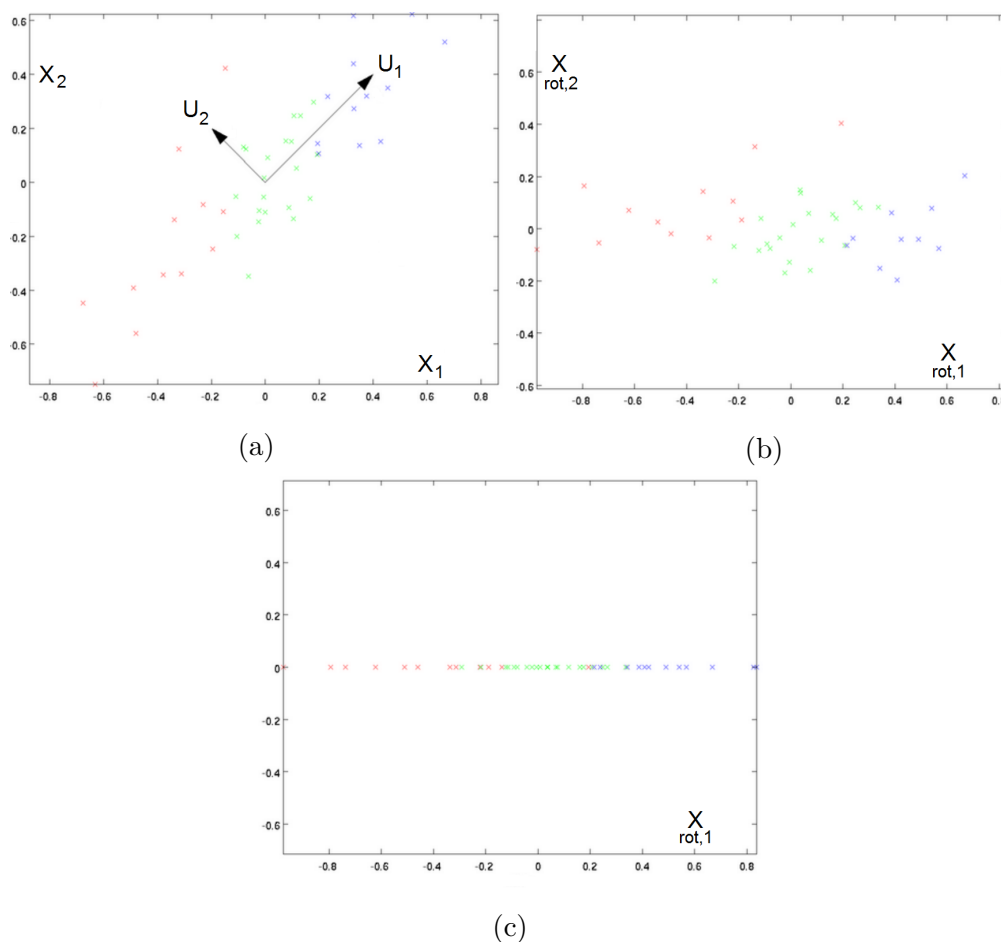
$$W_{proj} = \Lambda^{-1/2}U = \Sigma^{-1}U, \quad (3.21)$$

kjer je  $W_{proj}$  končna projekcijska matrika. Vsak vektor značilk  $y$  iz množice znanih ali neznanih slik je nato projiciran v nov vektorski podprostor po enačbi 3.22:

$$\tilde{y} = W_{proj}^T(y - \bar{x}). \quad (3.22)$$

Na opisan način dobimo novo predstavitev naših slik, kjer je vsak vektor značilk iz množice znanih in neznanih obrazov opisan z utežmi kot linearna kombinacija lastnih vektorjev učnih slik. Opisan proces lahko opazujemo na sliki 3.17. Nov vektorski prostor je manjši in bolj diskriminanten, kar

priporočam k večji natančnosti, predvsem pa k veliko hitrejšemu delovanju sistema.



Slika 3.17: (a) V hipotetičnem primeru imamo množico podatkov  $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$ , ki so opisani z dvema koordinatama v 2-D koordinatnem sistemu in velja  $x^{(i)} \in \mathbb{R}^2$ . Z razgradnjo singularnih vrednosti najprej poiščemo glavni komponenti  $U_1$  in  $U_2$ . (b) Če izračunamo  $x_{rot}^{(i)} = U^T x^{(i)}$  za vsak  $i$ , dobimo predstavitev podatkov  $x$  v novem koordinatnem sistemu, ki je opisan z glavnimi komponentami podatkov  $x$ . Pripis *rot* izhaja iz opazke, da opisan proces predstavlja rotacijo originalnih podatkov. (c) Če bi v drugem koraku ohranili le prvo glavno komponento ( $U_1$ ), ki izraža več variabilnosti kot druga, bi dobili takšno predstavitev podatkov.

### 3.5 Klasifikacija

V koraku klasifikacije se neznanim obrazom dodeli identiteta glede na podobnost, ki jo izražajo do znanih obrazov. Na področju prepoznavanja obrazov se v literaturi najpogosteje omenjajo metoda k najbližjih sosedov [40], metoda podpornih vektorjev [22], nevronske mreže [46] in LDA [56]. Pri našem delu smo zaradi računske hitrosti, hitre implementacije in dobrih rezultatov v literaturi [32, 33, 34] uporabili metodo k-najbližjih sosedov.

Za dani vhodni primer metoda k-najbližjih sosedov v množici učnih primerov (vektorjev) poišče  $k$  najbližjih oz. najbolj podobnih primerov. To stori glede na podano funkcijo za izračun razdalje med primeri. V našem sistemu smo za ta namen uporabili tri funkcije za izračun razdalje in sicer:

- **Kvadratna Chi razdalja:** to razdaljo smo uporabili za iskanje optimalnih nastavitev parametrov pri ekstrakciji značilk. To smo namreč počeli brez postopka zmanjševanja dimenzionalnosti, kar pomeni, da so se pri klasifikaciji primerjali kar histogrami podpodročij slik po ekstrakciji značilk, za kar pa je ta razdalja najbolj primerna. Kvadratna Chi razdalja med dvema vektorjema  $X = [x_1 x_2 \dots x_M]$  in  $Y = [y_1 y_2 \dots y_M]$  se izračuna po enačbi 3.23 [33]:

$$dist_{chi}(X, Y) = \sum_{i=0}^M \frac{(x_i - y_i)^2}{x_i + y_i} \quad (3.23)$$

- **Kosinusna razdalja:** to razdaljo smo uporabili pri testiranju metod za ekstrakcijo značilk Eigenfaces, LBP in ELBP. Razdaljo smo izbrali empirično ob analizi rezultatov. Kosinusna razdalja med dvema vektorjema  $X$  in  $Y$  se izračuna po enačbi 3.24 [32]:

$$dist_{cos}(X, Y) = -\frac{X^T Y}{\|X\| \|Y\|} \quad (3.24)$$

- **Obtežena kotna razdalja:** to razdaljo smo uporabili pri testiranju metod za ekstrakcijo značilk LPQ ter LPOG. Tudi ta razdalja je bila izbrana empirično. Obtežena kotna razdalja med dvema vektorjema  $X = [x_1 x_2 \dots x_M]$  in  $Y = [y_1 y_2 \dots y_M]$  se izračuna po enačbi 3.25 [34]:

$$dist_{wang}(X, Y) = -\frac{\sum_{i=1}^N z_i x_i y_i}{\sqrt{\sum_{i=1}^N x_i^2 \sum_{i=1}^N y_i^2}}, z_i = \sqrt{1/\lambda_i}, \quad (3.25)$$

kjer  $\lambda_i$  predstavlja lastne vrednosti iz koraka zmanjševanja dimenzionalnosti s PCA.

V tem poglavju smo torej opisali metode in postopke, ki smo jih uporabili v naši rešitvi. Naš sistem najprej normalizira sliko dane neznane osebe (rotacija, skaliranje, obrezovanje, svetlobna izravnava, frontalizacija, glajenje), ekstrahira značilke z izbrano metodo (Eigenfaces, LBP, ELBP, LPQ, LPOG), zmanjša dimenzionalnost vektorja značilk (PCA) ter osebi podeli identito glede na podobnost, ki jo izraža do slik znanih oseb (k-najbližjih sosedov), ki so bile pridobljene predhodno po istem postopku. V naslednjem poglavju sledi analiza testiranja našega sistema.



## Poglavje 4

# Analiza rezultatov

V tem poglavju bomo predstavili ter analizirali rezultate testiranja našega sistema na bazah FERET, SCface in ChokePoint v danem zaporedju. Rezultati bodo podani glede na različne metode za ekstrakcijo značilnk. Rezultate bomo ustrezno primerjali z najnovejšimi ter najboljšimi v literaturi. Nadalje bomo predstavili rezultate na frontaliziranih različicah baz FERET in SCface, pri slednji pa bomo opisali tudi vpliv glajenja robov. Prav tako bomo podali informacije o času, ki je potreben za delovanje našega sistema ter o nastavitvi parametrov metod za ekstrakcijo značilnk. Na koncu poglavja bomo strnili ugotovitve in za vsako metodo posebej argumentirali primernost za uporabo pri prepoznavanju obrazov pri nenadzorovanih pogojih.

### 4.1 Rezultati testiranja na bazi FERET

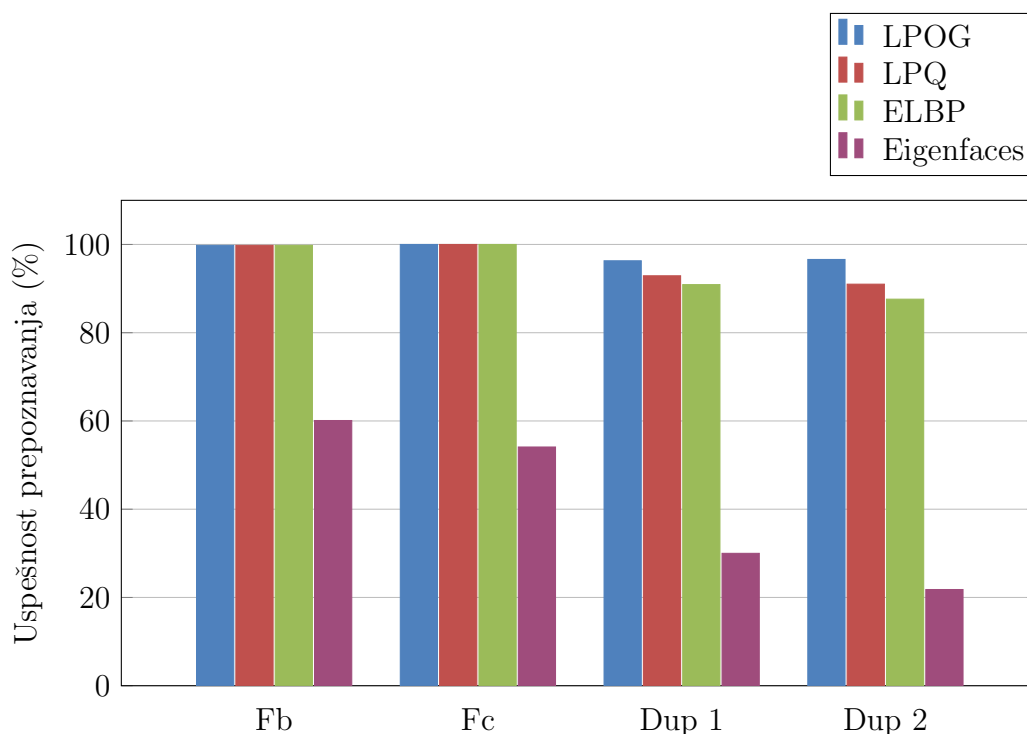
V tabeli 4.1 so podani rezultati testiranja baze FERET, ki je bilo izvedeno po standardnem protokolu [43]: posebej so se testirale množice slik Fb, Fc, Dup 1 ter Dup 2, za slike znanih obrazov pa je bila uporabljena množica Fa. Rezultati so primerjani s temi od avtorjev posamezne metode za ekstrakcijo značilnk ter z drugimi, ki so se v zadnjih letih izkazali kot konkurenčni v literaturi.

Tabela 4.1: Rezultati na bazi obrazov FERET. Naši rezultati so označeni z odebeljeno pisavo. Vsi rezultati so bili pridobljeni po standardnem protokolu, opisanem v Phillips et al. [43] in so urejeni po padajočem vrstnem redu glede na skupno povprečje.

Metoda	Fb	Fc	Dup1	Dup 2	Povprečje
Eigenfaces [42]	79.0	18.0	41.0	21.0	39.8
<b>Eigenfaces</b>	<b>60.1</b>	<b>54.1</b>	<b>30.0</b>	<b>21.8</b>	<b>41.5</b>
LBP [32]	98.7	99.0	83.9	78.2	90.0
<b>LBP</b>	<b>99.4</b>	<b>100</b>	<b>84.4</b>	<b>79.1</b>	<b>90.7</b>
ELBP [32]	99.4	100	89.1	86.8	93.8
<b>ELBP</b>	<b>99.8</b>	<b>100</b>	<b>90.9</b>	<b>87.6</b>	<b>94.6</b>
LPQ [33]	99.5	100	92.9	91.0	95.9
<b>LPQ</b>	<b>99.8</b>	<b>100</b>	<b>92.9</b>	<b>91.0</b>	<b>95.9</b>
EPOEM [31]	99.6	100	92.4	92.3	96.1
GOM [13]	99.8	100	95.7	93.1	97.2
SLF-RKR [69]	99.7	99.5	96.3	94.4	97.5
PLPQMC [33]	99.7	100	96.8	95.7	98.1
<b>LPOG</b>	<b>99.8</b>	<b>100</b>	<b>96.3</b>	<b>96.6</b>	<b>98.2</b>
LPOG [34]	99.8	100	97.4	97.0	98.6

Prva stvar, ki je razvidna iz tabele 4.1 je superiornost lokalnih metod za ekstrakcijo značilnk nad edino globalno metodo v tabeli, Eigenfaces, kar lahko opazujemo tudi na sliki 4.1. Priznati moramo, da se iskanju parametrov pri tej metodi nismo posvetili in so rezultati morda slabši, kot bi lahko bili, a s podajanjem rezultatov za metodo Eigenfaces smo le želeli dokazati prejšnjo trditev. Razliki med rezultati naše metode Eigenfaces ter rezultati avtorjev te metode [42] skoraj zagotovo botruje uporaba naprednejše rešitve za svetlobno izravnavo (Retina model), kar je najbolj očitno pri množici Fc, ki je namenjena testiranju sistemov pri nenadzorovani osvetlitvi.





Slika 4.1: Primerjava uspešnosti prepoznavanja na bazi FERET pri uporabi različnih metod za ekstrakcijo značilnk, podanih v legendi.

Če se sedaj posvetimo lokalnim metodam za ekstrakcijo značilnk je očitno, da so zelo odporne na spremembe pri izrazu ter osvetlitvi, testiranju česar sta namenjeni množici Fb in Fc v tem zaporedju. Naša rešitev je namreč prav pri vseh dosegla rezultat 99.8% ter 100%, kar je izenačen najboljši rezultat v literaturi. Primere pravilno klasificiranih obrazov iz množic Fb in Fc lahko opazujemo na sliki 4.2.

Večji problem predstavljata množici Dup 1 ter Dup 2, ki sta namenjeni prepoznavanju obrazov pri spremenljivi starosti. V tej množici so nekatere osebe slikane tudi več kot tri leta za tem, ko so bile pridobljene slike iz množice Fa, ki služijo kot slike znanih oseb. Vseeno se je tudi tu izkazala metoda LPOG, s katero smo skoraj dosegli najboljše rezultate v literaturi, o katerih poročajo avtorji metode v delu Nguyen et al. [34]. Podobne rezultate poročajo tudi v Caplier et al. (PLPQMC) [33], kjer so za ekstrakcijo značilnk



(a) Znani obrazi



(b) Neznani obrazi

Slika 4.2: Primeri pravilno klasificiranih obrazov iz množic Fb (na levi) in Fc (na desni).

uporabili Log-Gaborjeve filtre. Obetavne rezultate je metoda dosegla tudi na bazi SCface, o čemer se bomo prepričali v naslednjem razdelku. Dobre rezultate poročajo še v delih Zhang et al. (SLF-RKR) [69] ter Sun et al. (GOM) [13], a so se pri ekstrakciji značilk poslužili Gaborjevih valov, ki pa so približno štirikrat počasnejši od metode LPOG [31]. Ostale rešitve v literaturi ne poročajo bistveno boljših rezultatov od računsko zelo enostavnih metod ELBP ter LPQ. Pri tem je zanimivo opazovati še zelo dobre rezultate metode LPQ, posebno v primerjavi z deskriptorji LBP ter ELBP. Metoda je bila namreč izbrana zaradi odpornosti na zamegljenost (ang. *blurr*), ki pa v bazi FERET ni zastopana. Vseeno se je metoda izkazala boljše od LBP ter ELBP pri vseh zastopanih problematikah v bazi (spremenljivi izraz, osvetlitev in staranje).

Tabela 4.2: Rezultati na bazi obrazov FERET v primerjavi z njeno frontalizirano različico. Pripis F pomeni, da je množica frontalizirana.

Metoda	Fb	<b>Fb-F</b>	Fc	<b>Fc-F</b>	Dup1	<b>Dup1-F</b>	Dup2	<b>Dup2-F</b>
Eigenf.	60.1	<b>43.3</b>	54.1	<b>32.5</b>	30.0	<b>14.9</b>	21.8	<b>10.3</b>
LBP	99.4	<b>91.5</b>	100	<b>79.9</b>	84.4	<b>59.0</b>	79.1	<b>43.6</b>
ELBP	99.8	<b>92.1</b>	100	<b>78.9</b>	90.9	<b>62.1</b>	87.6	<b>51.3</b>
LPQ	99.8	<b>93.1</b>	100	<b>85.6</b>	92.9	<b>68.3</b>	91.0	<b>57.7</b>
LPOG	99.8	<b>94.9</b>	100	<b>91.2</b>	96.3	<b>75.3</b>	96.6	<b>67.1</b>

Našo rešitev smo stestirali tudi na frontalizirani različici baze FERET, kjer smo frontalizacijo opravili z metodo, ki je opisana v razdelku 3.2.2. Rezultati so podani v tabeli 4.2 ter na sliki 4.4. Zelo očitno je, da frontalizacija bistveno poslabša rezultate. Tako je npr. pri uporabi metode LPOG opazovan padec za kar 30% na frontalizirani različici množice Dup 2. Po temeljitnem pregledu slik smo ugotovili, da je razlog popačenje, do katerega pride pri frontalizaciji nekaterih obrazov, kar lahko opazujemo na sliki 4.3. Sicer so vse slike obrazov v bazi FERET pridobljene frontalno in je posledično niti ni smiselno frontalizirati. To smo storili zgolj zaradi primerjave rezultatov s frontalizacijo na bazi SCface, ki bo opisana v naslednjem razdelku.

Kakor je razvidno iz rezultatov podanih v tem razdelku je očitno, da baza FERET akademski skupnosti na področju prepoznavanja obrazov ne predstavlja več pravega izziva. Razlog tiči v tem, da so bile slike pridobljene pri nadzorovanih pogojih ter z visoko kvalitetnimi napravami. Množice Fb, Fc, Dup 1 ter Dup 2 so sicer namenjene raziskovanju prepoznavanja obrazov pri določenih problematikah, a se v vsaki množici posebej pojavlja le ena naenkrat. Množica Fb tako vsebuje le variacije pri izrazu, množica Fc izraža različno osvetlitev, množici Dup 1 ter Dup 2 pa spremenljivo starost. V praksi se pri veliki večini primerov pojavlja več problematik naenkrat, kar pa je dobro zastopano v bazi SCface, ki bo analizirana v naslednjem razdelku.

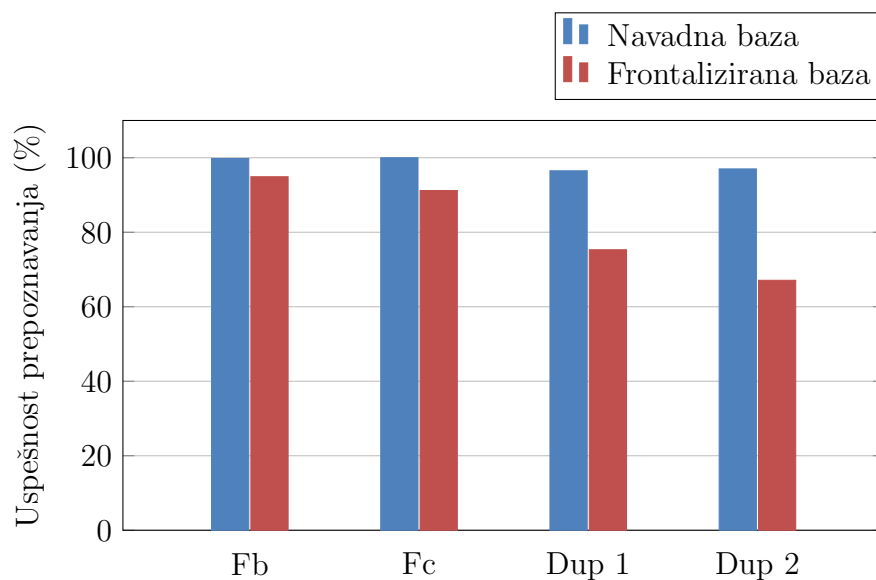


(a) Slike pred frontalizacijo.



(b) Slike po frontalizaciji.

Slika 4.3: Primeri uspešne (levo) ter neuspešne frontalizacije (desno).



Slika 4.4: Primerjava uspešnosti sistema na navadni FERET bazi ter na njeni frontalizirani različici ob uporabi metode LPOG za ekstrakcijo značilk.

V tem razdelku je potrebno omeniti še napaki pri poimenovanju nekaterih slik. V delu Nguyen et al. [31] so namreč opazili, da si dve osebi v množici  $F_a$  (znani obrazi) lastita vsak po dve slike, ki pa sta v obeh primerih drugačno poimenovani. Obrazi ter njihove identitete so prikazani na sliki 4.5. Posledica opisanega poimenovanja je, da naš sistem pri analiziranju množice  $F_b$  javi dve napaki, saj obrazoma z identitetama  $00277$  in  $00368$  podeli identiteti  $00185$  in  $00463$ . Če nas naš človeški mehanizem za prepoznavanje obrazov ne vara, bi si morali osebi z oznakama  $00185fa010$  in  $00463fa010$  deliti identiteto z osebama z oznakama  $00277fa010$  in  $00368fa010$ . V tem primeri bi naš sistem tudi pri množici  $F_b$  javil 100% uspešnost, a ker še nobeno delo v literaturi opisane napake ni dokumentiralo (z izjemo dela Nguyen et al. [31], kjer pa napake pri javljanju rezultatov prav tako niso upoštevali) smo tudi mi podali rezultate brez upoštevanja napake.



Slika 4.5: Dokumentacija napake pri poimenovanju slik v množici  $F_a$ .

## 4.2 Rezultati testiranja na bazi SCface

Testiranje na bazi SCface je potekalo po standardnem protokolu, opisanem v delu avtorjev baze Grgić et al. [19]. Pri tem se testiranje deli na t.i. dnevne (DayTime) in nočne (NightTime) teste. Pri prvih se primerjajo slike iz kamer 1 do 5 pri treh razdaljah s slikami znanih obrazov. Ta scenarij poda 15 možnih testnih množic, ki se razlikujejo tako v razdaljah do kamere kot v kakovosti kamer. Nočni testi so manj obsežni ter vsebujejo 6 testnih množic: slike iz kamer 6 in 7 pri treh razdaljah, pridobljene pri delovanju v infrardečem nočnem načinu. Pri tem se slike primerjajo z istimi obrazi kot pri dnevnih testih. Tabeli 4.3 in 4.4 vsebujeta rezultate testiranja naše rešitve na bazi SCface po opisanem protokolu.

Tabela 4.3: Rezultati na bazi SCface, pridobljeni po protokolu DayTime.

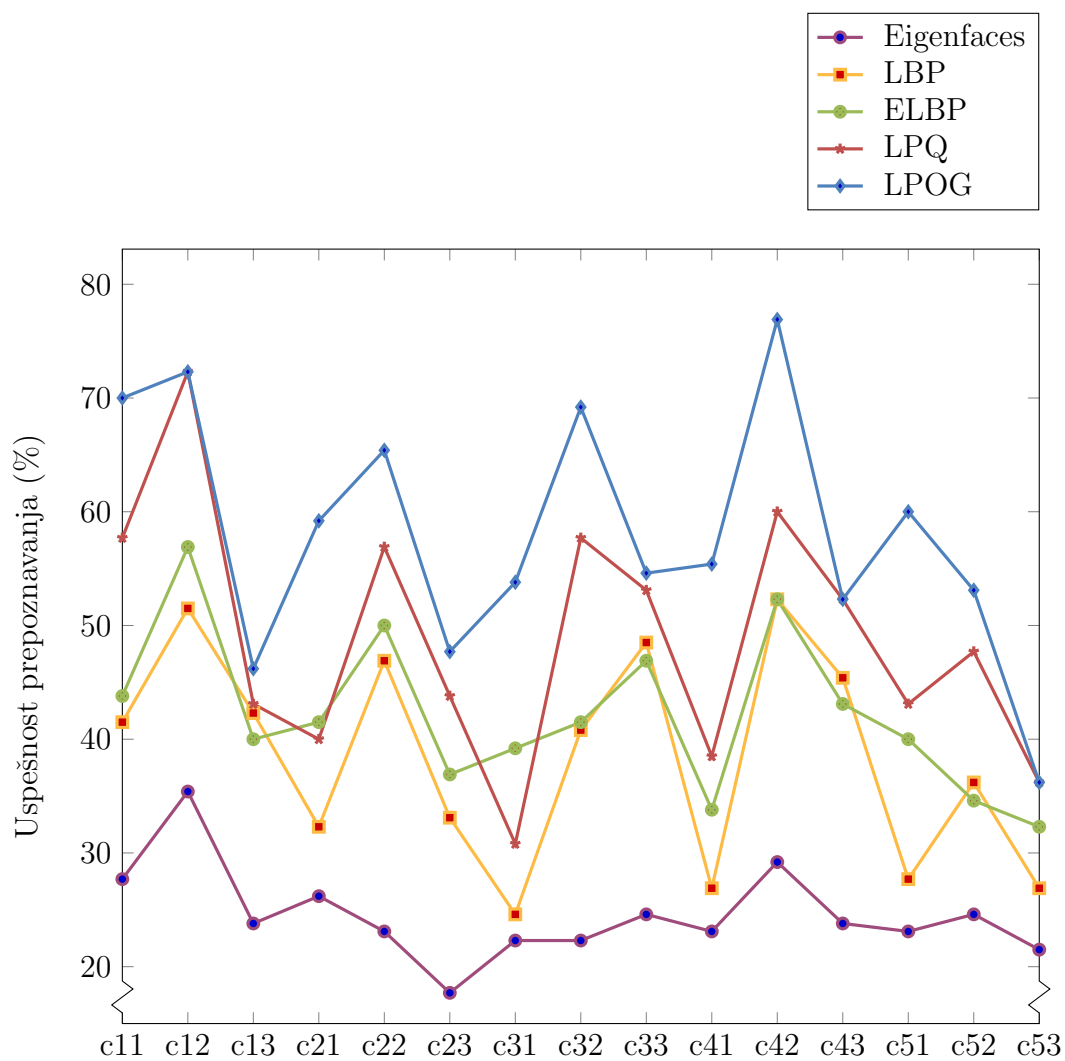
Kamera/Razdalja	Eigenfaces	LBP	ELBP	LPQ	LPOG
cam1_1	27.7	41.5	43.8	57.7	70.0
cam1_2	35.4	51.5	56.9	72.3	72.3
cam1_3	23.8	42.3	40.0	43.1	46.2
cam2_1	26.2	32.3	41.5	40.0	59.2
cam2_2	23.1	46.9	50.0	56.9	65.4
cam2_3	17.7	33.1	36.9	43.8	47.7
cam3_1	22.3	24.6	39.2	30.8	53.8
cam3_2	22.3	40.8	41.5	57.7	69.2
cam3_3	24.6	48.5	46.9	53.1	54.6
cam4_1	23.1	26.9	33.8	38.5	55.4
cam4_2	29.2	52.3	52.3	60.0	76.9
cam4_3	23.8	45.4	43.1	52.3	52.3
cam5_1	23.1	27.7	40.0	43.1	60
cam5_2	24.6	36.2	34.6	47.7	53.1
cam5_3	21.5	26.9	32.3	36.2	36.2
<b>Povprečje</b>	<b>24.6</b>	<b>38.5</b>	<b>42.2</b>	<b>48.9</b>	<b>58.2</b>

Najbolj očitna opazka iz tabele 4.3 je neuspešnost metod v primerjavi z bazo FERET, kar pa, sodeč po ogledu slik (na ogled v drugem poglavju na sliki 2.3) niti ne čudi. Slike namreč izražajo veliko variabilnosti pri osvetlitvi in pozi ter so hkrati zelo slabe ločljivosti. Ponovno je moč opazovati tudi superiornost lokalnih metod za ekstrakcijo značilnk nad globalno metodo Eigenfaces. Zanimivost, ki ni bila prisotna pri testiranju na bazi FERET so variacije rezultatov pri nekaterih testnih množicah ob uporabi različnih metod, kar je dobro razvidno tudi na sliki 4.7. Čeprav so povprečni rezultati različnih metod skladni z opazovanji pri bazi FERET, lahko opazimo, da so pri nekaterih množicah rezultati slabši pri metodi, ki dosega boljši povprečni rezultat. Tako je npr. rezultat pri množicah *cam1-3*, *cam3-3* in *cam4-3* boljši pri LBP kot pri ELBP. To je verjetno posledica dejstva, da so slike pri tretji razdalji najmanj frontalne ter pride prednost uporabe elips manj do izraza. Motivacija za uporabo eliptičnih vzorcev za ekstrakcijo pri metodi ELBP je namreč opazovanje, da imajo to obliko najbolj informativne značilke na obrazu (oči, usta, nos), ki pa so v tej skupini na nekaterih slikah zelo neizrazne. To je moč opazovati na sliki 4.6.



Slika 4.6: Primeri slik, pridobljenih pri razdalji 3.

Podobno so rezultati pri nekaterih množicah boljši pri uporabi ELBP kot pri LPQ, čeprav slednja dosega bistveno boljši povprečni rezultat. To se zgodi pri množicah *cam2-1* ter *cam3-1*, ki sta pridobljeni pri največji razdalji. To pomeni, da so obrazi zelo frontalni, kjer pa se metoda ELBP dobro obnese. Na podlagi boljših rezultatov pri krajših razdaljah lahko sklepamo tudi, da je metoda LPQ v primerjavi z ELBP ter LBP zelo odporna na spremenljivo pozo.



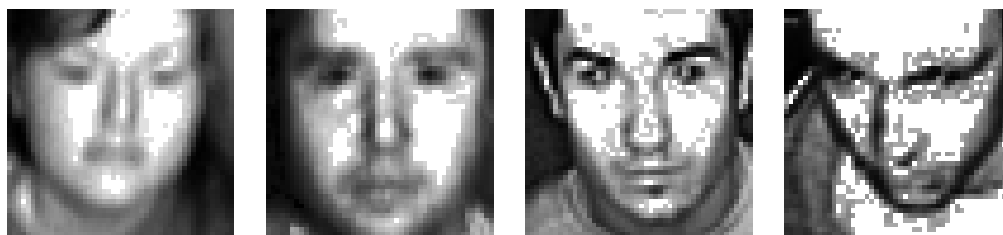
Slika 4.7: Primerjava uspešnosti prepoznavanja obrazov na bazi SCface pri uporabi različnih metod za ekstrakcijo značilk, podanih v legendi. Oznake na horizontalni osi predstavljajo kamere in razdalje (c11 predstavlja kamero 1 pri razdalji 1). Rezultati so podani le za test DayTime.



Še zadnja, a zelo očitna opazka je premoč metode LPOG v primerjavi s preostalimi, kar je skladno z opazovanji na bazi FERET. Dodatno je iz slike 4.7 razvidno, da zadevna metoda dosega najboljše rezultate pri prav vseh množicah dnevnih testov. To ne čudi, saj metoda združuje prednosti deskriptorjev ELBP ter LPQ. Drugače je pri nočnih (NightTime) testih, kjer so obrazne lastnosti zelo slabo izražene. To je posledica delovanja kamer v infrardečem nočnem načinu, ki zelo slabo ohranja prepoznavnost obraznih značilk, kar je dobro razvidno iz slike 4.8. Tako je v tabeli 4.4 moč opazovati ekstremen padec uspešnosti pri vseh metodah, iz česar lahko zaključimo, da kljub naprednim metodam za normalizacijo in ekstrakcijo značilk problema nenadzorovane osvetlitve še zdaleč ne moremo označiti kot rešenega.

Tabela 4.4: Rezultati na bazi SCface, pridobljeni po protokolu NightTime.

Kamera/Razdalja	Eigenfaces	LBP	ELBP	LPQ	LPOG
cam6_1	7.7	10.7	12.3	10.0	16.9
cam6_2	13.8	18.5	16.9	16.2	23.8
cam6_3	13.1	17.7	21.5	22.3	28.5
cam7_1	10.8	14.6	13.8	12.3	17.7
cam7_2	15.4	15.4	17.7	20.8	20.0
cam7_3	16.2	14.6	15.4	19.2	18.5
<b>Povprečje</b>	<b>12.8</b>	<b>15.3</b>	<b>16.2</b>	<b>16.8</b>	<b>21.3</b>



Slika 4.8: Primeri slik, prisotnih pri nočnem (NightTime) testu.

V tabeli 4.5 so naši rezultati primerjani z najboljšimi v literaturi. Razvidno je, da smo, vsaj kolikor nam je znano, na bazi SCface z metodo LPOG dosegli **najboljše rezultate v literaturi**. Pri tem je potrebno dodati, da smo bili deležni precejšnje pomoči s strani avtorjev dela Nguyen et al. [34], še posebej na področju normalizacije slik. Pri bazi SCface se rezultati namreč občutno izboljšajo, če pri različnih razdaljah ohranimo različno število pikslov med očmi. S tem se glavne značilke bolje poravnajo s tistimi na slikah znanih obrazov. Glavni razlog za izboljšanje rezultatov iz del [32, 33, 34] (LBP, ELBP, LPQ, LPOG) tiči v koraku glajenja robov, ki smo ga dodali pri obdelavi slik. Opazili smo namreč, da so se zaradi majhnega števila pikslov pri postopkih obdelave začele pojavljati nezaželenje anomalije v obliki nazobčanih robov, šuma ter črt. S korakom glajenja, opisanega v razdelku 3.2.5, smo zmanjšali vpliv omenjenih anomalij, ki so sicer bile s prostim očesom komaj vidne.

Tabela 4.5: Povprečni rezultati za protokola DayTime in NightTime v primerjavi z ostalimi rešitvami v literaturi. Naši rezultati so označeni z odebeljeno pisavo.

Metoda	DayTime	NightTime
Eigenfaces-Fa [42]	4.7	3.2
<b>Eigenfaces</b>	<b>24.6</b>	<b>12.8</b>
LBP [32]	37.2	11.8
<b>LBP</b>	<b>38.5</b>	<b>15.3</b>
ELBP [32]	42.7	15.0
<b>ELBP</b>	<b>42.2</b>	<b>16.2</b>
EPOEM [31]	43.9	15.4
LPQ [33]	46.6	16.2
<b>LPQ</b>	<b>49.1</b>	<b>17.6</b>
PLPQMC [33]	55.3	20.8
LPOG [34]	56.8	20.9
<b>LPOG</b>	<b>58.2</b>	<b>21.3</b>

Vsi preostali rezultati v literaturi na bazi SCface so bili pridobljeni z metodami, ki so že bile povzete v prejšnjem razdelku pri analizi rezultatov na bazi FERET. Posebno opombo si zaslužijo najslabši rezultati v tabeli, kjer lahko opazimo pripis *Fa*, kar pomeni, da je bila za učenje algoritma PCA uporabljena različna množica od tiste, ki se je uporabila kot referenca za znane obraze. V tem primeru je bila to množica *Fa* iz baze FERET. Avtorji dela Grgić et al. [19], ki so hkrati tudi lastniki baze SCface, zagovarjajo uporabo različne množice za učenje. To naj bi bil edini način za pošteno primerjavo različnih rezultatov, saj se le-ti bolje približajo realni situaciji. V tem primeru se namreč sistem namesti v neznano okolje, ki uporablja drugačno množico slik od teh, na katerih je bil naučen. S to motivacijo smo tudi v našem delu objavili rezultate, pri katerih smo metodo PCA naučili na množici *Fa* baze FERET. Rezultate je moč opazovati v tabeli 4.6. Zanimivo pri tem je, da se pri nočnih testih v dveh primerih rezultati na ta način celo izboljšajo. To je verjetno posledica dejstva, da je pri našem sistemu po uporabi metode PCA vsaka slika opisana z linearno kombinacijo slik znanih obrazov, ki jih je le 130. Le-te imajo zelo malo skupnega s slikami nočnih testov, ki jih je moč opazovati na sliki 4.8. Množica *Fa* po drugi strani vsebuje 1,196 slik, kar daje sistemu veliko širšo paleto kombinacij za opis posamezne slike.

Tabela 4.6: Povprečni rezultati za protokola DayTime in NightTime v primerjavi z rezultati, kjer je za učenje algoritma PCA bila uporabljena množica *Fa* baze FERET in ne množica znanih obrazov SCface, kar je dodatno označeno s pripisom *Fa*.

Metoda	DayTime	<b>DayTime-Fa</b>	NightTime	<b>NightTime-Fa</b>
Eigenfaces	24.6	<b>19.7</b>	12.8	<b>14.7</b>
LBP	38.5	<b>34.2</b>	15.3	<b>12.8</b>
ELBP	42.2	<b>39.2</b>	16.2	<b>17.4</b>
LPQ	49.1	<b>40.3</b>	17.6	<b>13.7</b>
LPOG	58.2	<b>48.1</b>	21.3	<b>17.6</b>

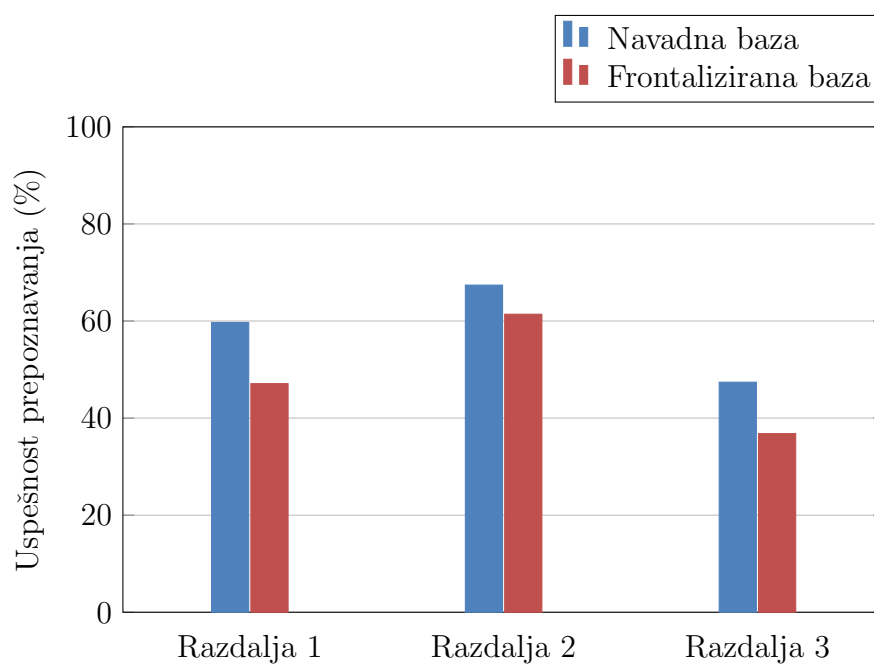
Na isti način kot pri bazi FERET smo frontalizirali tudi bazo SCface. Kot je razvidno iz tabele 4.7 in slike 4.9 se tudi tokrat rezultati niso izboljšali, a je pri le-teh možno opazovati manjši padec, kot pri bazi FERET. Pri uporabi globalne metode Eigenfaces je celo moč opazovati porast rezultatov pri razdalji 2. Rezultate smo tokrat podali le glede na različno razdaljo, kar je drugače kot je to običajno pri uporabi standardnega protokola. Motivacija za to je opazka, da poza obrazov najbolj variira s spreminjanjem razdalje, kar je posledica spreminjajočega se kota med kamero in osebo.

Tabela 4.7: Povprečni rezultati za vsako razdaljo na bazi obrazov SCface v primerjavi z njeno frontalizirano različico, ki je označena s pripisom F. Pri rezultatih niso upoštevane kamere v nočnem načinu.

Razdalja	Eigen.	<b>Eigen.-F</b>	ELBP	<b>ELBP-F</b>	LPOG	<b>LPOG-F</b>
1 (4,2m)	24.5	<b>21.5</b>	39.7.0	<b>35.5</b>	59.7	<b>47.1</b>
2 (2,6m)	26.9	<b>28.8</b>	47.1	<b>45.8</b>	67.4	<b>61.4</b>
3 (1,0m)	22.3	<b>18.9</b>	39.8.0	<b>30.5</b>	47.4	<b>36.8</b>
Povprečje	24.6	<b>23.1</b>	42.2	<b>37.3</b>	58.2	<b>48.4</b>

Glavni razlog, da je razkorak med rezultati tukaj manjši kot pri bazi FERET je dobra sposobnost poravnanja obrazov pri metodi za frontalizacijo. Kot smo opisali v razdelku 3.2, smo pri rotiranju in skaliranju vseh baz uporabili ročno pridobljene koordinate značilk. Pri bazi FERET so le-te pridobljene učinkovito in so slike posledično dobro poravnane. Drugače je pri bazi SCface kjer so slike majhne ter nizke ločljivosti, zaradi česar ročna lokalizacija značilk predstavlja večji problem. To je razvidno po sami obdelavi slik, kjer je očitno, da slike ponekod niso najboljše poravnane. To je razvidno iz slike 4.10, kjer so predstavljene tudi frontalizirane slike istih oseb.

Kot je bilo moč zaznati v tem razdelku, predstavlja baza SCface akademski skupnosti na področju prepoznavanja obrazov še vedno zelo velik izziv. V bazi je namreč zastopana velika večina problematik, s katerimi se sistemi soočajo pri prepoznavanju obrazov na javnih mestih. Podobne lastnosti vsebujejo tudi slike v bazi ChokePoint, katere analiza sledi.



Slika 4.9: Primerjava uspešnosti sistema na navadni SCface bazi ter na njeni frontalizirani različici ob uporabi metode LPOG za ekstrakcijo značilk.



(a) Slike pred frontalizacijo.



(b) Slike po frontalizaciji.

Slika 4.10: Primeri uspešne (levo) ter neuspešne frontalizacije (desno).

### 4.3 Rezultati testiranja na bazi ChokePoint

Baza obrazov ChokePoint je nastala leta 2011 pod vodstvom avtorjev dela Chen et al. [64]. Namen baze je testiranje rešitev za prepoznavanje obrazov po načinu video-to-video, kar pomeni, da se v koraku klasifikacije poljubno število sličic neznane osebe primerja z  $n$  množicami sličic znanih oseb poljubne velikosti, saj vsak posnetek vsebuje poljubno veliko število sličic. Da smo bazo priredili za testiranje na našem sistemu, ki deluje po načinu image-to-image, smo sličice vsakega posnetka povprečili v eno samo sliko, kakor je to opisano v razdelku 2.3. Tako smo pridobili 16 različnih testnih množic, ki vsebujejo do 30 obrazov in ki hkrati igrajo tudi vlogo množic znanih obrazov, kakor to določa protokol, opisan v delu Chen et al. [64]. Ta namreč zahteva, da se omenjenih 16 testnih množic najprej razdeli v dve skupini po 8 (G1 in G2), kjer nato vsaka množica izmenoma zavzame vlogo množice znanih obrazov, med tem, ko se preostalih 7 množic uporabi za testiranje. Ta postopek se torej izvede dvakrat, po enkrat za vsako izmed dveh skupin. Razdelitev množic v dve skupini je podana v tabeli 4.8.

Tabela 4.8: Razdelitev množic v dve skupini G1 in G2.

G1	P1E_S1_C1	P1E_S2_C2	P2E_S2_C2	P2E_S1_C3
	P1L_S1_C1	P1L_S2_C2	P2L_S2_C2	P2L_S1_C1
G2	P1E_S3_C3	P1E_S4_C1	P2E_S4_C2	P2E_S3_C1
	P1L_S3_C3	P1L_S4_C1	P2L_S4_C2	P2L_S3_C3

Rezultati testiranja na bazi ChokePoint so podani v tabeli 4.9, kjer so primerjani z rezultati iz literature. Pridobili smo jih po zgoraj opisanem protokolu, kjer smo izvedli 112 testov (dve skupini po 8 testnih množic, kjer vsaka enkrat zavzame vlogo množice znanih obrazov, preostalih 7 pa se uporabi za testiranje, torej  $2 * 8 * 7 = 112$ ) in podali povprečen rezultat.

Opazimo lahko, da so rezultati različnih metod v skladu z opazovanji na bazah FERET in SCface. Tudi tokrat je edina globalna metoda Eigenfaces dosegla najslabši rezultat. Presenetljivo je, da je razlika med omenjeno metodo ter preostalimi lokalnimi metodami manjša, kot pri bazah

Tabela 4.9: Rezultati na bazi obrazov ChokePoint. Naši rezultati so označeni z odebeljeno pisavo. Pripis CP pomeni, da smo algoritem PCA učili na slikah, ki so služile tudi kot referenca za znane obraze, pripis Fa pa pomeni, da smo algoritem PCA učili na množici Fa baze FERET.

Metoda	Rezultat-CP	Rezultat-Fa
<b>Eigenfaces</b>	<b>61.8</b>	<b>78.2</b>
Asym_shrp [64]	75.4	/
DFFS [64]	83.4	/
Gabor_asym [64]	84.0	/
<b>LBP</b>	<b>84.7</b>	<b>81.7</b>
<b>ELBP</b>	<b>85.5</b>	<b>84.9</b>
PPIQA [64]	86.7	/
<b>LPQ</b>	<b>86.9</b>	<b>86.5</b>
<b>LPOG</b>	<b>92.5</b>	<b>92.3</b>

SCface in FERET. To gre pripisati majhnosti množic znanih obrazov, ki pri bazi ChokePoint v povprečju vsebujejo zgolj 27 obrazov. Sklepamo lahko torej, da se uspešnost sistemov za prepoznavanje obrazov povečuje z zmanjševanjem števila znanih obrazov. To je sicer logično, saj primerki v manjših množicah v povprečju izražajo veliko večjo razdaljo do najbližjega sosedu, kot v številčnejših množicah.

Nadalje je iz rezultatov razvidno, da smo tudi tokrat, vsaj kolikor nam je znano, z metodo LPOG dosegli **najboljše rezultate v literaturi** na bazi ChokePoint. Dodati je potrebno, da smo edine rezultate, ki upoštevajo standarden protokol našli le v delu Chen et al. [64], kjer pa se niso posvečali optimizaciji sistema za prepoznavanje obrazov, pač pa algoritmom za izbiro najbolj primernih sličic iz enega posnetka. To so storili tako, da so vsaki sličici podali štiri ocene, ki odražajo kvaliteto določene lastnosti slike. Izbrane lastnosti so poravnane in poza obraza ter ostrina in osvetlitev slike.

Po pridobljenih ocenah so nato izbrali  $N$  najbolj primernih sličic neznanih obrazov za primerjavo z  $N$  sličicami znanih obrazov. Pri tem so ugotovili, da je optimalno število sličic 16. Za ekstrakcijo značilnk so uporabili metodo multi-regijskih histogramov (ang. *Multi-region Histograms*) [48] ter metodo skupnega podprostora (ang. *Mutual Subspace Method*) [66] za klasifikacijo. Kljub preprosti klasifikaciji z metodo  $k$ -najbližjih sosedov se je naša rešitev pri uporabi deskriptorjev LPQ in LPOG obnesla bolje.

Zaradi razlogov, ki smo jih opisali že pri analizi rezultatov na bazi SCface, smo tudi tokrat podali rezultate pri uporabi drugačne množice za učenje algoritma PCA, kar smo ponovno storili z uporabo množice Fa baze FERET. Rezultate pridobljene na ta način lahko prav tako opazujemo v tabeli 4.9. Opazimo lahko, da je pri delovanju v tem načinu sistem dosegel skoraj tako dobre rezultate, kakor takrat, ko smo za učenje uporabili slike znanih oseb, pri metodi Eigenfaces pa celo veliko boljše. Podobno kot pri rezultatih na bazi SCface pri nočnih testih je verjetno tudi tukaj razlog za dobre rezultate dejstvo, da je znanih slik obrazov v povprečju le 27. Če sistem učimo na množici Fa baze FERET, ima tako metoda PCA na voljo veliko več možnih kombinacij za opis posamezne slike, kakor če to storimo na množici znanih oseb.

Ker smo že na bazah FERET in SCface pokazali, da uporabljena rešitev za frontalizacijo ne izboljša rezultatov, je pri bazi ChokePoint nismo stestirali. Tudi drugače smo se bazi ChokePoint zaradi izraženih lastnosti slik, ki so zelo podobne tem v bazi SCface, manj posvetili. Menimo, da če bi baza vsebovala podobno število slik znanih obrazov, kakor baza SCface, bi si bili rezultati verjetno zelo podobni.

## 4.4 Nastavitev parametrov

Pri testiranju posamezne baze smo uporabljali empirično določene parametre, ki so podali najboljši rezultat in jih je moč razbrati iz tabel 4.10, 4.11, 4.12 in 4.13. Takšna nastavitev sicer ni najbolj zanesljiva, saj je sistem na ta način



optimiziran za čim večjo robustnost proti problematikam, ki nastopajo le v posamezni bazi. Argumentiramo, da je v realni situaciji še najbolj zanesljiva nastavitev parametrov pri bazi SCface, ker le-ta najbolj posnema pogoje, ki so prisotni pri prepoznavanju obrazov na javnih mestih. Če torej pri metodi LPOG uporabimo nastavitev parametrov za bazo SCface na vseh treh bazah (le pri bazi FERET ohranimo število pod-regij), dosežemo rezultate, ki so prikazani na sliki 4.11. Razvidno je, da s tako nastavitvijo parametrov naš sistem še vedno dosega odlične rezultate, ki so le malenkostno slabši od empirično nastavljenih nastavitev. Opravičljivo bi bilo s spreminjanjem velikosti slik spreminjati tudi radije ter velikost okna (M), a se rezultati niso bistveno spremenili. Podrobnejši opis parametrov je podan pri opisu metod za ekstrakcijo značilk v razdelku 3.3.

Tabela 4.10: Nastavitve parametrov za metodo LBP

Parameter	FERET	SCface	ChokePoint
Radij (R)	5	3	3
Sosedni (N)	8	7	11
Pod-regije	$9 \times 9$	$6 \times 6$	$9 \times 9$

Tabela 4.11: Nastavitve parametrov za metodo ELBP

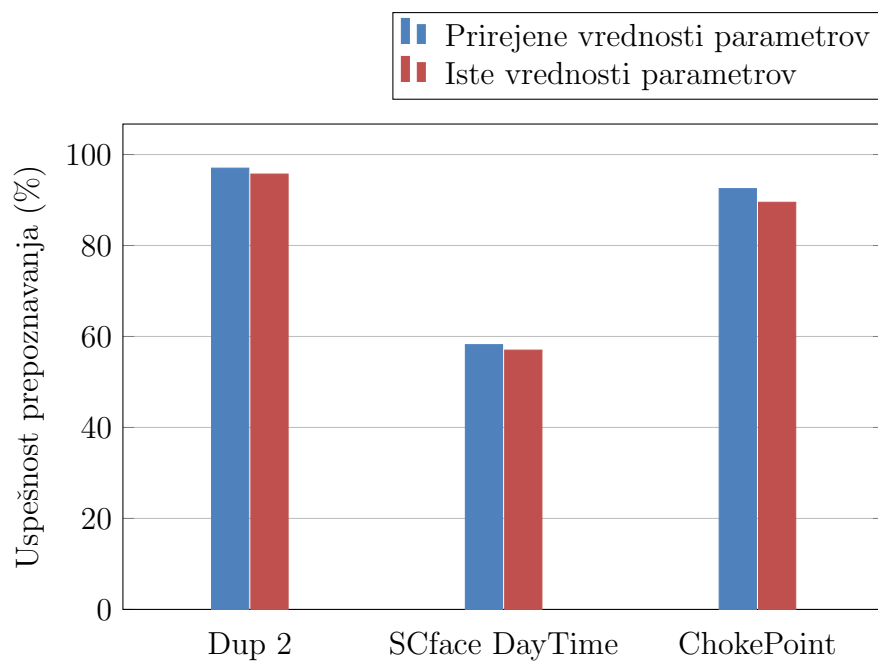
Parameter	FERET	SCface	ChokePoint
Radij-h (R1)	5	5	3
Radij-v (R2)	3	3	2
Sosedni (N)	8	7	11
Pod-regije	$9 \times 9$	$6 \times 6$	$9 \times 9$

Tabela 4.12: Nastavitve parametrov za metodo LPQ

Parameter	FERET	SCface	ChokePoint
Velikost okna (M)	9	9	7
Pod-regije	$10 \times 10$	$5 \times 5$	$9 \times 9$

Tabela 4.13: Nastavitve parametrov za metodo LPOG

Parameter	FERET	SCface	ChokePoint
Radij-h (R1)	5	5	3
Radij-v (R2)	3	3	2
Sosedi (N)	8	11	11
Velikost okna (M)	11	11	9
BELBP pod-regije	$9 \times 9$	$4 \times 4$	$6 \times 6$
LPQ pod-regije	$10 \times 10$	$4 \times 4$	$6 \times 6$



Slika 4.11: Primerjava uspešnosti sistema na podanih množicah ob uporabi metode LPOG pri prirejenih parametrih ter pri parametrih, ki so enaki za vse tri baze.

V sklopu nastavitve parametrov je potrebno omeniti še število ohranjenih glavnih komponent v projekcijski matriki. Pri našem delu smo pri posamezni bazi rezultate vedno pridobili tako, da smo v projekcijski matriki ohranili od 75% do 100% glavnih komponent in zabeležili najboljše rezultate. Tega postopka so se poslužili tudi avtorji uporabljenih metod. Vseeno smo to področje premalo raziskali, da bi si upali iz podatkov karkoli sklepati. Opazili smo namreč razkorak pri številu ohranjenih glavnih komponent pri različnih bazah. Kljub temu resnega problema to ne predstavlja, saj pri ohranitvi različnega števila glavnih komponent (v območju med 75% in 100%) rezultati niso bistveno oscilirali in se niso nikoli poslabšali za več kot 2%.

## 4.5 Časovna zahtevnost

Pri sistemih za prepoznavanje obrazov, ki delujejo v realnem času, je časovni strošek skoraj tako pomemben kakor učinkovitost. S to motivacijo smo v tabeli 4.14 podali informacije o času, ki je potreben za ekstrakcijo značilk ter kreacijo projekcijske matrike, ki sta najbolj časovno zahtevni operaciji v našem sistemu. Informacije o času smo pridobili tako, da smo z našim sistemom 50-krat zgradili projekcijsko matriko iz množice Fa baze FERET, kar vključuje tudi ekstrakcijo značilk vseh slik. Postopek smo opravili dvakrat: enkrat s slikami velikosti  $128 \times 128$  pikslov ter drugič s slikami velikosti  $48 \times 48$  pikslov.

Iz tabele je razvidno, da metoda Eigenfaces po hitrosti močno prekaša ostale metode za ekstrakcijo značilk, vseeno pa so se dobro izkazale tudi vse preostale metode. V našem delu sicer ne poročamo o hitrosti popularnejših rešitev za ekstrakcijo značilk, so pa to storili avtorji metode LPOG v delu Nguyen et al. [31], kjer je omenjena metoda delo opravila 4-krat hitreje kakor v literaturi zelo priljubljene rešitve, ki so osnovane na Gaborjevih valovih. V delu so poročali o hitrosti ekstrakcije 44-ih slik na sekundo, v kolikor smo mi dosegli v povprečju le 23 slik na sekundo, a smo pri našem delu razpolagali z manj zmogljivo strojno opremo. Kot rečeno smo v tabeli 4.14 podali tudi

Tabela 4.14: Čas, potreben za ekstrakcijo značilnk ter celoten čas, ki zajema tudi kreacijo projekcijske matrike iz množice Fa (1196 slik) baze FERET.

Metoda	Velikost slike	Celoten čas [s]	Čas e. značilnk [ms]	Slike/s
Eigenf.	$128 \times 128$	4.68	1.19	838
LBP	$128 \times 128$	10.31	7.62	131
ELBP	$128 \times 128$	18.47	13.92	71
LPQ	$128 \times 128$	17.92	11.34	88
LPOG	$128 \times 128$	61.85	42.57	23
Eigenf.	$48 \times 48$	1.83	0.79	1261
LBP	$48 \times 48$	4.41	2.92	341
ELBP	$48 \times 48$	6.46	4.55	220
LPQ	$48 \times 48$	6.74	4.47	223
LPOG	$48 \times 48$	20.75	15.47	64

informacije o času, ki je potreben za kreacijo projekcijske matrike, a to niti ne bi bilo potrebno, saj se omenjena operacija opravi malokrat. Če sistem svoje algoritme uči na množici znanih obrazov, je to potrebno opraviti le takrat, ko v sistem vnesemo novo sliko znanega obraza. V nasprotnem primeru, torej če sistem učimo na drugačni množici, se operacija kreiranja projekcijske matrike opravi le enkrat. Na ta način bo uspešnost sistema manjša, a v primerih, kjer so sistemu konstantno predstavljeni novi znani obrazi, je to lahko edini način.

Pri našem delu smo vse teste opravili na prenosnem računalniku Lenovo E540 (CPE i7-4702MQ @ 2,2 GHz, 4Gb RAM) z operacijskim sistemom Windows 10 (64 bit) in v okolju Matlab R2016b (64 bit).

## 4.6 Povzetek analize rezultatov

V tem poglavju smo predstavili rezultate, ki smo jih z našim sistemom dosegli na bazah obrazov FERET, SCface in ChokePoint. Pri bazi FERET, ki je najbolj raziskana baza obrazov v akademskem svetu, smo se zelo približali najboljšim rezultatom v literaturi, med tem, ko smo jih pri bazah SCface in ChokePoint presegli. Ugotovili smo, da je metoda za ekstrakcijo značilnk LPOG robustna proti večini problematik, s katerimi se sistemi za prepoznavanje obrazov soočajo pri delovanju ob nenadzorovanih pogojih. Nadalje naši rezultati nakazujejo na to, da je slike, ki so majhne in nizke ločljivosti priporočljivo zgladiti, kakor smo to mi storili na bazi SCface. Rešitev, ki je na začetku veliko obetala, a se je izkazala za še premalo raziskano je frontalizacija. Slednjo smo namreč uporabili na bazah FERET in SCface, a so se pri obeh rezultati poslabšali. Vseeno argumentiramo, da si metoda zasluži nadaljnih raziskav, saj smo pri nekaterih podmnožicah opazili porast uspešnosti. Naš sistem je hitrejši od popularnih rešitev v literaturi, s čimer je zagotovo primeren za razvoj izven meja prototipa, kar smo dodatno dokazali tudi z učenjem sistema na množici obrazov, ki je različna od množice znanih obrazov. Trditev podpiramo tudi z dobrimi rezultati pri testiranju sistema pri isti nastavitvi parametrov za vse tri baze. Kljub temu, kakor je rezvidno iz rezultatov na bazi SCface, bo potrebnega še veliko dela, da bodo sistemi zmožni učinkovitega delovanja na mestih, kjer hkrati nastopa več zelo izraženih problematik.



## Poglavje 5

### Zaključek

Magistrsko delo opisuje sistem za prepoznavanje obrazov pri nenadzorovanih pogojih, razvoj katerega se je začel s študijem področja, nadaljeval z identifikacijo najboljših tehnologij in metod, končal pa z implementacijo prototipa in s testiranjem. Glede na omenjene podprocese smo na grobo tudi strukturirali nalogo. V prvem sklopu smo tako opisali področje ter problematike, s katerimi se sooča in argumentirali, da je problem prepoznavanja obrazov na javnih mestih še zelo aktualen. V nadaljevanju smo opisali baze obrazov (FERET [43], SCface [19], ChokePoint [64]), ki smo jih uporabili pri testiranju, za tem pa temeljiteje predstavili posamezne module našega sistema. Nalogo smo zaključili z analizo testiranja, v sklopu katerega smo pri bazah SCface in ChokePoint, vsaj kolikor nam je znano, celo dosegli najboljše rezultate v literaturi. V tem poglavju smo tudi podprli trditve o robustnosti našega sistema s podajanjem dobrih rezultatov pri uporabi posebne testne množice obrazov.

Če se natančneje dotaknemo prispevkov naloge, bi na tem mestu izpostavili natančen opis sistema, s katerim je moč doseči zelo dobre rezultate pri različnih pogojih. Pri tem smo ugotovili da se pri majhnih slikah nizke ločljivosti rezultati izboljšajo, če jih v koraku obdelave slik zgladimo. Še pred začetkom testiranja smo si veliko obetali od nove metode za frontalizacijo obrazov, opisane v delu Hassner et al. [21], a se rezultati niso izboljšali.

Tekom izdelave smo na vseh treh bazah preizkusili tudi kopico različnih fotometričnih metod za svetlobno izravnavo, ki so agregirane v knjižnici INface, a sta se najboljše odrezali uveljavljeni metodi Retina model na bazah FERET in ChokePoint ter histogramska izravnavo na bazi SCface. Kljub temu smo na dveh bazah z metodo ekstrakcije značilk LPOG [34] izboljšali rezultate v literaturi.

V raziskavo smo se spustili z motivacijo, da sistem za prepoznavanje obrazov implementiramo na javnem mestu. Ta motivacija je še vedno prisotna, za kar pa sistemu manjka še par komponent, začevši z modulom za avtomatsko zaznavo značilk. Če sistem s to tehnologijo ne razpolaga, je potrebna ročna lokalizacija značilk, kar pa je zamudno in drago. Na srečo na tem področju obstajajo številne zanesljive rešitve, med drugimi Viola in Jones [62], Rowley [46] in ERT [24]. Nadalje je pri prepoznavanju na javnih mestih najverjetnejši scenarij video-to-image, kar pomeni, da neznano osebo predstavlja video in ne slika, kakor je to primer pri trenutnem stanju našega sistema. Da bi le-tega priredili omenjenemu scenariju, je sistem potrebno nadgraditi v koraku klasifikacije. Ena izmed metod, ki jo je pri tem možno uporabiti, je metoda skupnega podprostora (ang. *Mutual Subspace Method*) [66], katero so uporabili avtorji baze ChokePoint (namen te baze je ravno v testiranju pri scenariju video-to-image ali video-to-video) v svojem delu Chen et al. [64].

V kolikor je potrebno določene komponente sistemu do delovanja na javnem mestu še dodati, pa je moč obstoječe module še izboljšati z raziskavo na naslednjih področjih:

- **Uporaba barvnih slik:** pri našem delu smo prepoznavanje obrazov vedno raziskovali le na črno-belih slikah. V delih [57, 27, 16] so pri tem upoštevali tudi barvo.
- **Določitev kvalitete slik:** dejali smo še, da je video-to-image najbolj verjeten scenarij pri prepoznavanju obrazov v resničnem svetu. V tem primeru imamo na voljo več kot le eno sliko neznane osebe, s čimer se odpre več možnosti za njihovo uporabo. Ena izmed možnosti je primerjanje vseh slik naenkrat s sliko znanega obraza, kar smo že opisali



v prejšnjem odstavku. Naslednja možnost je, da iz posnetka izberemo le  $N$  najbolj kvalitetnih slik za primerjavo, za kar pa je potreben algoritem za določitev kvalitete slik. Številne rešitve so opisane v delih [64, 14, 1].

- **Fuzija slik:** še ena možnost, ki je na voljo pri razreševanju video-to-image scenarija je fuzija slik. Nekaj podobnega smo uporabili tudi sami pri bazi ChokePoint, ko smo povprečili vse sličice enega posnetka, da smo pridobili eno sliko. Naprednejše metode, ki opravljajo to nalogo so opisane v delih [73, 6].
- **Super-ločljivost:** pri bazi ChokePoint in še posebej pri bazi SCface smo imeli opravka s slikami zelo nizke ločljivosti. Z izjemo tega, da smo pri bazi SCface uporabili glajenje robov, smo slike tretirali na isti način kot slike pri bazi FERET, ki so visoke ločljivosti. V delih [51, 65] pri reševanju tega problema uporabljajo metode t.i. super-ločljivosti.

Na tem mestu naj omenimo, da so pri nekaterih izmed zgoraj omenjenih del svoje rešitve testirali tudi na bazah FERET, SCface in ChokePoint, a tega niso počeli po standardnem protokolu ali pa so celo uporabljali slike iz različnih podatkovnih baz hkrati. Dodati je potrebno še, da v kolikor se rezultati pri upoštevanju zgoraj naštetih rešitev lahko izboljšajo (lahko pa tudi ne), je še vedno potrebno v obzir vzeti hitrost sistema, ki pa je za nas zelo pomembna in se pri dodajanju rešitev ustrezno zmanjša.

Izmed v uvodu naštetih problematik smo pri našem (v večjem ali manjšem obsegu) stestirali vse razen prekrivanja. Pri slednji imamo v mislih predvsem večje predmete kot so npr. sončna očala, pokrivala, šali, rute itd. V skupen koš lahko vržemo tudi nekatere lastnosti, ki obrazu pripadajo intrinzično (npr. brade, brki), a jih pri našem delu nismo testirali ločeno ter posledično o uspešnosti naše rešitve pri teh pogojih ne moramo poročati. Razlog za to je predvsem nedostopnost testnih podatkov. Vseeno smo dokazali, da je naš sistem konkurenčen z rešitvami v literaturi in robusten na široko paleto problematik. Optimistični smo, da bomo v bližnji prihodnosti našo rešitev še nadgradili ter jo uspešno razširili izven meja prototipa.



# Literatura

- [1] A. A. Abayomi-Alli, E. O. Omidiora, S. O. Olabiyisi, and J. A. Ojo. Adaptive regression splines models for predicting facial image verification and quality assessment scores. *Balkan Journal of Electrical and Computer Engineering*, 3(1):17–26, 2015.
- [2] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *European conference on computer vision*, pages 469–481. Springer, 2004.
- [3] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 28(12):2037–2041, 2006.
- [4] T. Ahonen, E. Rahtu, V. Ojansivu, and J. Heikkila. Recognition of blurred faces using local phase quantization. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pages 1–4. IEEE, 2008.
- [5] A. Ammato, A. Lanata, and E. Scilingo. Comparative study on photometric normalization algorithms for an innovative, robust and real-time eye gaze tracker. *Journal of Real-Time Image Processing*, 8(1):21–33, 2011.
- [6] L. An, B. Bhanu, and S. Yang. Face recognition in multi-camera surveillance videos. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2885–2888. IEEE, 2012.

- [7] J. Anil and S. Pankanti. Biometrics systems: anatomy of performance. *IEICE transactions on information and systems*, 84(7):788–799, 2001.
- [8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. In *European Conference on Computer Vision*, pages 43–58. Springer, 1996.
- [9] J. R. Beveridge, P. J. Phillips, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, et al. The challenge of face recognition from digital point-and-shoot cameras. In *IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2013.
- [10] H. S. Bhatt, R. Singh, and M. Vatsa. Covariates of face recognition. *Technical Reports from DSpace at IIIT-Delhi*, 2015.
- [11] S. G. Bhele and V. H. Mankar. A review paper on face recognition techniques. *International Journal of Advanced Research in Computer Engineering and Technology*, 1(8):339–346, 2012.
- [12] Brussels airport attack. <http://www.dailymail.co.uk/news/article-3507492/A-blood-spattered-baby-girl-screams-dead-mother-Disturbing-footage-emerges-child-crying-amidst-wreckage-Brussels-airport-bombing.html>. Dostop: 2016-08-17.
- [13] Z. Chai, Z. Sun, H. Vazquez-Mendez, R. He, and T. Tan. Gabor ordinal measures for face recognition. *IEEE Transactions on Information Forensics and Security*, 9(1):14–26, 2014.
- [14] J. Chen, Y. Deng, G. Bai, and G. Su. Face image quality assessment based on learning to rank. *IEEE Signal Processing Letters*, 22(1):90–94, 2015.
- [15] J.-C. Chen, S. Sankaranarayanan, V. M. Patel, and R. Chellappa. Unconstrained face verification using fisher vectors computed from fronta-

- lized faces. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–8. IEEE, 2015.
- [16] J. Y. Choi, Y. M. Ro, and K. N. Plataniotis. Color local texture features for color face recognition. *IEEE Transactions on Image Processing*, 21(3):1366–1380, 2012.
- [17] J. G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America. A, Optics and image science*, 2(7):1160–1169, 1985.
- [18] M. De Marsico, M. Nappi, D. Riccio, and H. Wechsler. Robust face recognition for uncontrolled pose and illumination changes. *IEEE Transactions on Systems*, 43(1):149–163, 2013.
- [19] M. Grgić, K. Delac, and S. Grgić. Scface – surveillance cameras face database. *Multimedia Tools and Applications Journal*, 51(3):863–879, 2011.
- [20] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*, 47:23–34, 2016.
- [21] T. Hassner, S. Harel, E. Paz, and R. Enbar. Effective face frontalization in unconstrained images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4295–4304, 2015.
- [22] B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: Global versus component-based approach. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 688–694. IEEE, 2001.
- [23] H. Imtiaz and S. A. Fattah. A face recognition scheme using wavelet based dominant features. *Signal and Image Processing: An International Journal*, 2011.

- [24] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. *Computer Vision and Pattern Recognition*, pages 1867–1874, 2014.
- [25] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(1):103–108, 1990.
- [26] E. Land and J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.
- [27] S. H. Lee, J. Y. Choi, Y. M. Ro, and K. N. Plataniotis. Local color vector binary patterns from multichannel face images for face recognition. *IEEE Transactions on Image Processing*, 21(4):2347–2353, 2012.
- [28] A. M. Martinez. The AR face database. *CVC Technical Report*, 24, 1998.
- [29] A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2001.
- [30] S. P. Mudunuri and S. Biswas. Low resolution face recognition across variations in pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(6):978–992, 2015.
- [31] H. T. Nguyen. *Contributions to Facial Feature Extraction for Face Recognition*. PhD thesis, University of Grenoble, 2014.
- [32] H. T. Nguyen and A. Caplier. Elliptical local binary patterns for face recognition. *Computer Vision*, pages 85–96, 2012.
- [33] H. T. Nguyen and A. Caplier. Patch-based local phase quantization of monogenic components for face recognition. *International Conference on Image Processing*, pages 229–233, 2014.

- [34] H. T. Nguyen and A. Caplier. Local patterns of gradients for face recognition. *IEEE Transactions on Information Forensics and Security*, 10(8):1739–1751, 2015.
- [35] M. J. Nordin and A. A. K. A. Hamid. Combining local binary pattern and principal component analysis on t-zone face area for face recognition. *Pattern Analysis and Intelligent Robotics*, pages 25–30, 2011.
- [36] M. A. Ochoa-Villegas, J. A. Nolasco-Flores, O. Barron-Cano, and I. A. Kakadiaris. Addressing the illumination challenge in two-dimensional face recognition: a survey. *IET Computer Vision*, 38(5):1034–1040, 2016.
- [37] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [38] V. Ojansivu and J. Heikkila. Blur insensitive texture classification using local phase quantization. *Image and Signal Processing*, pages 236–243, 2008.
- [39] D. N. Parmar and B. B. Mehta. Face recognition methods and applications. *International Journal of Computer Technology and Applications*, 4(1):84–86, 2013.
- [40] P. Parveen and B. Thuraisingham. Face recognition using multiple classifiers. *2006 18th IEEE International Conference on Tools with Artificial Intelligence*, pages 179–186, 2006.
- [41] V. Perlibakas. Distance measures for pca-based face recognition. *Pattern Recognition Letters*, 25(6):711–724, 2004.
- [42] H. Phillips, P. J. ands Moon, S. A. Rizvi, and P. J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

- 
- [43] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
  - [44] S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. Romeny, J. B. Zimmerman, and K. Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.
  - [45] R. Ravnik and F. Solina. Audience measurement of digital signage: Quantitative study in real-world environment using computer vision. *Interacting With Computers*, 25(3):218–228, 2013.
  - [46] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.
  - [47] J. Sadr, I. Jarudi, and P. Sinha. The role of eyebrows in face recognition. *Perception*, 32(3):285–293, 2003.
  - [48] C. Sanderson and B. C. Lovell. Multi-region probabilistic histograms for robust and scalable identity inference. *Lecture Notes in Computer Science*, 5558:199–208, 2009.
  - [49] M. Schatten, M. Bača, and M. Čubrilo. Towards a general definition of biometric systems. *IJCSI International Journal of Computer Science Issues*, 2, 2009.
  - [50] H. Schneiderman and T. Kanade. Object detection using the statistics of parts. *International Journal of Computer Vision*, 56(3):151–177, 2004.
  - [51] C. Senthilsingh and M. Manikandan. A novel algorithm for face recognition from very low resolution images. *Journal of Electrical Engineering and Technology*, 10(2):659–669, 2015.



- [52] P. Sinha, B. Balas, O. Y., and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.
- [53] K. K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):39–51, 1998.
- [54] Y. Taigman, M. Yang, M. Renzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1701–1708, 2014.
- [55] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Transactions on Image Processing*, 19(6):1635–1650, 2010.
- [56] S. Thakur, J. K. Sing, D. K. Basu, and M. Nasipuri. Face recognition using fisher linear discriminant analysis and support vector machine. *Contemporary Computing*, 40:318–326, 2009.
- [57] P. Tome, R. Vera-Rodriguet, J. Fierrez, and J. Garcia-Ortega. Fusion of facial regions using color information in a forensic scenario. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, 8259:399–406, 2013.
- [58] V. Štruc and N. Pavešič. *Photometric Normalization Techniques for Illumination Invariance*. IGI Global, 2011.
- [59] V. Štruc and N. Pavešič. Photometric normalization techniques for illumination invariance. *Advances in Face Image Analysis: Techniques and Technologies*, pages 279–300, 2011.
- [60] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

- 
- [61] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.
  - [62] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 1:511–518, 2001.
  - [63] N.-S. Vu and A. Caplier. Illumination-robust face recognition using retina modeling. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 3289–3292. IEEE, 2009.
  - [64] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. *IEEE Biometrics Workshop, Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 81–88, 2011.
  - [65] X. Xu, W. Liu, and L. Li. Low resolution face recognition in surveillance systems. *Journal of Computer and Communications*, 2:70–77, 2014.
  - [66] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. *Automatic Face and Gesture Recognition*, pages 318–323, 1998.
  - [67] W. S. Yambor, B. A. Draper, and J. R. Beveridge. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. *Empirical Evaluation Methods in Computer Vision*, 2002.
  - [68] K. Yan, Y. Chen, and D. Zhang. Gabor surface feature for face recognition. *Asian Conference on Pattern Recognition*, pages 288–292, 2011.
  - [69] M. Yang, Z. L., S. Shiu, and D. Zhang. Robust kernel representation with statistical local features for face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 24(6):900–912, 2013.

- 
- [70] M. H. Yang. Face detection. *Encyclopedia of Biometrics*, pages 447–452, 2015.
  - [71] M. H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
  - [72] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
  - [73] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi. Recent advances on singlemodal and multimodal face recognition: A survey. *IEEE Transactions on Human-Machine Systems*, 44(6):701–714, 2015.